THE UNIVERSITY *of* EDINBURGH

Kinetic Monte Carlo simulations of autocatalytic protein aggregation

Kym Eden

Doctor of Philosophy
The University of Edinburgh
2014

Success only opens up new avenues of brilliant possibility.
Such a task is eternal and its joys without bounds.

— Graham McNeill, A Thousand Sons [68]

Summary

Biology provides many examples of materials that are able to arrange themselves into complicated, ordered structures. **Protein** molecules, for instance, form most of the machinery within a biological cell, but in order to perform their function they must first "fold" into a compact form and then, in many cases, assemble into larger structures containing multiple proteins. This ability to **self-assemble** is of great interest for the development of new materials; for example, the formation of silk inspired the creation of Nylon. There are, however, situations where the self-assembly of certain molecules can cause problems for an organism; for example, the presence of **amyloid fibrils** has been linked with Alzheimer's, Huntington's and CreutzfeldtJakob diseases, among many others.

Amyloid fibrils are long chains formed from protein molecules, and may appear stiff (like dry spaghetti) or flexible (like hair). Each fibril grows by the addition of more protein molecules to its ends. It is theoretically possible for any protein to assemble into a fibril, although some do so more readily than others. These chains may be many thousands of proteins long, but the typical thickness of an amyloid fibril is only $^{1}/_{10,000}{}^{\text{th}}$ that of a human hair.

It is difficult to analyse how amyloid fibrils self-assemble due to their small size. This is particularly true in the early stages of self-assembly because the amyloid fibrils will be few in number and very short. Thus, in order to study them it is necessary to create models of the self-assembly process. The predications provided by these models can then be compared to the effects which can be seen in experiments.

In one of the most recent models, the self-assembly of proteins into amyloid fibrils is accelerated by the **autocatalytic** creation of extra fibrils by the fragmentation of existing ones. Autocatalysis is a form of positive feedback, whereby new material is created in proportion to the existing material. In this case, fibril ends are created when an existing fibril breaks. This happens in proportion to the total length of all fibrils; the more fibril ends that exist, the faster the protein can aggregate. Since this model involves the breaking of fibrils into smaller pieces, it is expected to result in the formation of a large number of short fibrils once they are no longer able to grow. This is concerning because current theories indicate that short fibrils may cause cell damage that leads to the onset of disease, while longer fibrils are inert.

This "fragmentation model" is the one that I will be mainly discussing in this thesis. Using it, I have developed predictions for the early stages of amyloid fibril self-assembly which indicate that some experimentally observable effects will depend on the sample volume. This means that some of the effects which are observed in experiments conducted in volumes typically used in the laboratory, may change or vanish when similar experiments are conduced in a smaller volume comparable with a biological cell.

I have also explored the separate effects of modifications to this model. In the first of these I prevent fibrils from breaking once a critical threshold concentration of fibrils is reached. This modified model appears to explain experimentally observed effects, and also suggests that altering the environmental conditions (such as adding salt to the sample) can reduce the number of short fibrils produced. The other modification to the model allows fibril ends to join together. This also reduces the number of short fibrils and can lead to the formation of loops. Such loops have been observed experimentally, and this model offers a first glimpse at understanding their importance to amyloid fibril self-assembly.

Combined, my findings show that changes in the environment — for example by adding salt or changing the volume of the container — could be expected to fundamentally alter which mechanisms are important for amyloid fibril self-assembly.

Abstract

The self-assembly of proteins into filamentous structures underpins many aspects of biology, from dynamic cell scaffolding proteins such as actin, to the amyloid plaques responsible for a number of degenerative diseases. Typically, these self-assembly processes have been treated as nucleated, reversible polymerisation reactions, where dynamic fluctuations in a population of monomers eventually overcome an energy barrier, forming a stable aggregate that can then grow and shrink by the addition and loss of more protein from its ends.

The nucleated, reversible polymerisation framework is very successful in describing a variety of protein systems such as the cell scaffolds actin and tubulin, and the aggregation of haemoglobin. Historically, amyloid fibrils were also thought to be described by this model, but measurements of their aggregation kinetics failed to match the model's predictions. Instead, recent work indicates that autocatalytic polymerisation — a process by which the number of growth competent species is increased through secondary nucleation, in proportion to the amount already present — is better at describing their formation. In this thesis, I will extend the predictions made in this mean-field, autocatalytic polymerisation model through use of kinetic Monte Carlo simulations.

The ubiquitous sigmoid-like growth curve of amyloid fibril formation often possesses a notable quiescent lag phase which has been variously attributed to primary and secondary nucleation processes. Substantial variability in the length of this lag phase is often seen in replicate experimental growth curves, and naively may be attributed to fluctuations in one or both of these nucleation processes. By comparing analytic waiting-time distributions, to those produced by kinetic Monte Carlo simulation of the processes thought to be involved, I will demonstrate

that this cannot be the case in sample volumes comparable with typical laboratory experiments.

Experimentally, the length of the lag phase, or "lag time", is often found to scale with the total protein concentration, according to a power law with exponent γ. The models of nucleated polymerisation and autocatalytic polymerisation predict different values for this scaling exponent, and these are sometimes used to identify which of the models best describes a given protein system. I show that this approach is likely to result in a misidentification of the dominant mechanisms under conditions where the lag phase is dominated by a different process to the rest of the growth curve. Furthermore, I demonstrate that a change of the dominant mechanism associated with total protein concentration will produce "kinks" in the scaling of lag time with total protein concentration, and that these may be used to greater effect in identifying the dominant mechanisms from experimental kinetic data.

Experimental data for bovine insulin aggregation, which is well described by the autocatalytic polymerisation model for low total protein concentrations, displays an intriguing departure from the predicted behaviour at higher protein concentrations. Additionally, the protein concentration at which the transition occurs, appears to be affected by the presence of salt. Coincident with this, an apparent change in the fibril structure indicates that different aggregation mechanisms may operate at different total protein concentrations. I demonstrate that a transition whereby the self-assembly mechanisms change once a critical concentration of fibrils or fibrillar protein is reached, can explain the observed behaviour and that this predicts a substantially higher abundance of shorter filaments — which are thought to be pathogenic — at lower total protein concentrations than if self-assembly were consistently autocatalytic at all protein concentration.

Amyloid-like loops have been observed in electron and atomic-force microscographs, together with non-looped fibrils, for a number of different proteins including ovalbumin. This implies that fibrils formed of these proteins are able to grow by fibrillar end-joining, and not only monomer addition as is more commonly assumed. I develop a simple analytic expression for polymerisation by monomer addition and fibrillar end-joining, (without autocatalysis) and show that this is not sufficient to explain the growth curves obtained experimentally for ovalbumin.

I then demonstrate that the same data can be explained by combining fibrillar end-joining and fragmentation. Through the use of an analytic expression, I estimate the kinetic rates from the experimental growth curves and, via simulation, investigate the distribution of filament and loop lengths.

Together, my findings demonstrate the relative importance of different molecular mechanisms in amyloid fibril formation, how these might be affected by various environmental parameters, and characteristic behaviour by which their involvement might be detected experimentally.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Parts of this work have been published in Morris et al. [74].

(*Kym Eden*)

Acknowledgements

There are many people whom I wish to thank for their help and support during the work represented by this thesis. In particular:

My supervisor, Cait MacPhee, for her guidance, patience, and enthusiasm; especially while I was writing this thesis.

Rosalind Allen, for her "it's all coming together" motivation, technical assistance, and meticulous proof-reading of my work.

Fellow postgraduate student Ryan Morris, for performing most of the experimental work to which I refer herein, and whose collaboration helped shape development of the models.

Jason Kalapothakis and Jay Gillam, for providing other experimental results and valuable group discussions.

Juraj Szavitz-Nossan, for finding analytic solutions to the models indicated in the text, and bringing new insight and enthusiasm to the project.

The Edinburgh Compute and Data Facilities, for providing the processing power necessary to perform large numbers of repeat simulations.

The Engineering and Physical Sciences Research Council for their financial support.

And of course, my postgraduate peers, with whom I shared an office and department for four years (give or take), for providing valued input and consultation, or a welcome distraction. Of these people, Robert Concannon and Alastair Mailer deserve particular mention for providing extensive assistance on probability theory and data analysis.

Finally, I wish to thank my partner, Holly Bontoft, and all of my friends and family for their support, and for taking an interest in my work.

I dedicate this thesis to all those whose memories have faded; that they may never fade from ours.

Contents

List of Figures

Chapter 1

Introduction

1.1 An overview of polymers

Polymers are a diverse class of macromolecules, consisting of long chains of smaller molecules referred to as monomers; the class encompasses a wide variety of natural and synthetic materials. A polymer may consist entirely of one type of monomer (homopolymers), or a mixture of different monomers (copolymers). In either case the chemical properties of the monomers determine the polymer's flexibility, solubility, strength, and other mechanical and chemical properties; in many cases the scaling of the polymer's physical properties with chain length can be explained by simple theories, independent of the chemical nature of the monomers [22].

This thesis will focus on a group of biological polymers known as "amyloid fibrils". However, microscopic detail of the monomers and mechanisms will be largely neglected, and many of the results are expected to hold for polymer self-assembly processes in general.

Typically, polymers grow longer by the addition of monomers to one or both of their ends in a process called polymerisation, but some can also grow when two individual strands of polymer join, end-to-end. Depolymerisation is the process by which a polymer looses material from its ends, but in most cases the monomers are covalently bound within the polymer and this is not a significant process.

Synthetic polymers have become a familiar part of modern industry and medicine due to their diverse properties and easy customisation. Plastics such as polythene, polystyrene and nylon (the group of polyamides) form the packaging, casing and gearing of numerous machines and devices, while other polymers such as polyepoxides (epoxy) and polyvinyl acetate (PVA) form the core component of resins and glues. In the constant search to design new materials, researchers have increasingly turning to biology for inspiration.

Biochemistry is heavily reliant on polymers: the instructions necessary for a biological cell to function are preserved and propagated in the ordered sequence of nucleotides forming a deoxyribonucleic acid (DNA) molecule; while the arrangement of sugars in a polysaccharide can provide either an energy store or structural support. Of the myriad bio-polymers, perhaps the most remarkable group are the polypeptides, which include proteins [(1)]. Proteins form the core machinery for the entire biological cell. They control and facilitate the activation of genes, mediate most intracellular and extracellular reactions, form sensors, scaffolding, motors, and more. Together, proteins constitute more than half of a cell's dry mass [77].

Polypeptides are built from amino acids and are typically copolymers. Many varieties of amino acid exist but life typically only utilises a small subset [77]. All share a common chemical backbone but with different sidechain group which can be used to categorise amino acids by their preferred chemical environment. Some are non-polar, and often strongly hydrophobic as a consequence, while the rest are either charged or able to be polarised by neighbouring charge, and prefer an aqueous environment.

In order to minimise the free energy of the copolymer as a whole, a polypeptide tends to fold itself into a configuration that reduces the exposure of hydrophobic groups to surrounding water while also reducing exposure of hydrophilic groups to surrounding lipids [86]. This is thought to be the primary driver for what is typically referred to as "protein folding", whereby the polypeptide assembles itself into a compact structure. As it becomes more compact, however, charge distribution becomes important too, either aiding or preventing certain parts of the polypeptide from coming into close proximity. These electrostatic

(1) Conventionally, polypeptide is used to refer to any amino acid polymer, while protein is reserved for those with biological function.

interactions, along with hydrogen and sulphur bonding, also help to stabilise the preferred folded state — often referred to as the native conformation — once it has been found [21].

Many proteins assemble into small complexes to perform a function. These contain several folded protein molecules held together by hydrophobic or electrostatic interactions. However, some proteins self-assemble into much larger complexes: polymers in their own right, with individual proteins as their monomers. Two examples are actin and tubulin which form the filaments and microtubules that function as both cell structure and highways along which motile proteins pull their cargo. These protein filaments are dynamic and are constantly being assembled and dismantled, as required. Other polymer-forming proteins, for example silk, cannot be readily dismantled and are immensely strong. While these particular examples of proteins self-assembling into larger structures are desirable, in other cases this kind of aggregation can cause problems for the cell.

The cellular environment is extremely crowded and many reactions take place in close proximity, simultaneously. Alterations to the amino acid sequence due to point mutations in the genetic information, can affect the native conformation of a protein molecule [77]. In most cases these substitutions either have no apparent effect or render the protein completely dysfunctional, potentially leading to dysfunction or death of the cell [2]. However, in some cases alterations the amino acid sequence only affects the folding process, leading to a higher propensity for unfolding, partial folding, or missfolding [20, 21]. Under these circumstances, proteins in a crowded environment are likely to aggregate due to the exposure of their hydrophobic regions, to one another. Temperature, pH and other environmental factors also have an impact on the strength of attraction between certain side chains and hence the stability of the protein, leading to similar problems. This makes protein aggregation a limiting factor to the range of environmental conditions over which a particular organism can live.

Most protein aggregates are disordered clumps of misfolded proteins which form in the cell near the sites of protein synthesis; here, the newly formed proteins may not have had sufficient time to find their native conformation before bumping into one another. Most cells have mechanisms for coping with these aggregates and in a healthy cell they will be quickly broken up. However, not all protein aggregates are disordered or easily removed.

Amyloid fibrils are a group of highly ordered and extremely resilient protein aggregates [21]. They have been implicated in a number of degenerative disorders such as Alzheimer's disease and Creutzfeldt-Jakob disease, but are also increasingly being found to be functionally employed by organisms, owing to their structural strength and ability to self-assemble. Amyloid fibrils are another example of a polymer formed of proteins, and for the remainder of the thesis the word "monomer" will be used to refer to the individual proteins within an amyloid-like fibril.

(a) Schematic of the cross-β structure of amyloid fibrils, by Nelson and Eisenberg [78].

(b) Schematic of the arrangement of peptides in a β-sheet configuration, by Caporini et al. [8].

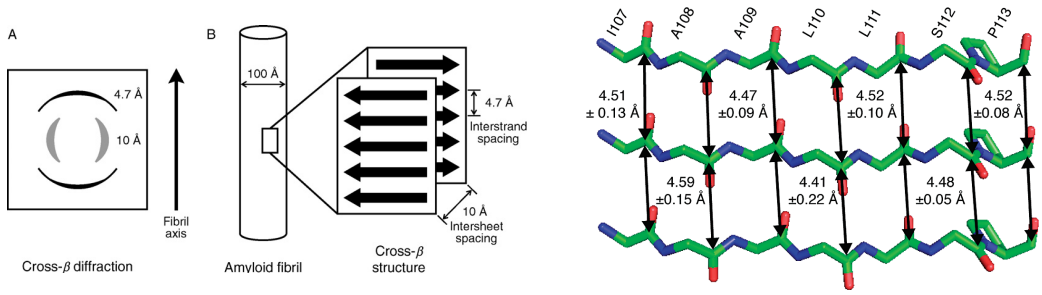


Figure 1.1 Illustrations of the β-sheet structure of amyloid fibrils. The schematic in **(aA)** shows the X-ray diffraction pattern produced by this cross-β structure, with meridional and equatorial reflections indicating spacings of ~ 4.7 Å and ~ 10 Å, respectively for this fibril. These are also indicated in **(aB)** as the interstrand and intersheet spacings, respectively, where the strands are individual polypeptide chains, and the sheets are individual amyloid filaments. The fibril itself is made of several of these filaments and has a diameter of ~ 100 Å. These dimensions vary between proteins, although the interstrand spacing is constrained by the range of a hydrogen bond [8, 23, 78, 99]; a schematic representation of the hydrogen bond network between the peptide backbones of adjacent peptide strands is shown in **(b)**, using the protein $\text{TTR}_{105-115}$ [8] for illustration purposes.

Amyloid fibrils have a high aspect ratio, typically with lengths of several micrometres and widths of only a few tens of nanometres. They are characterised by a so-called cross β-sheet structure (illustrated in figure 1.1a), which is a common motif found in the native conformation of many proteins and provides a great deal of structural strength [77]. In this structure, polypeptide chains are aligned with their backbones parallel to one another as illustrated in figure 1.1b.

The peptide backbone of every polypeptide consists of alternating amine and carboxyl groups which possess very slightly different charge distributions. When two sections of peptide backbone align, the difference in charge distribution creates a network of hydrogen bonds. While these bonds are individually weak, a large number of them can provide a high degree of resistance to tensile and sheer stress. This ubiquitous β-sheet structure means that any polypeptide has the potential to form amyloid fibrils which, once formed, are very difficult for the cell to remove.

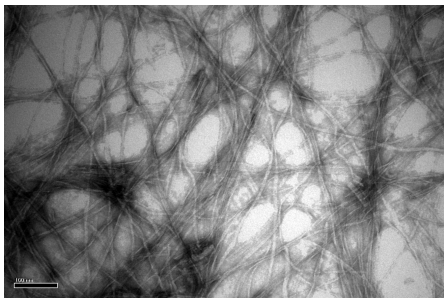
While the presence of amyloid fibrils has been linked with neurodegenerative diseases, increasing evidence indicates that it is not the mature fibrils, but rather the smaller, precursor aggregates that cause the damage; however, the reason for their toxicity remains unknown [7]. Given that any protein can adopt the amyloid β-sheet structure, the plethora of diseases in which they have been implicated is unsurprising. However, most of these are associated with later life — even those with a known genetic link — which suggests that the initial steps in the formation of amyloid fibrils have a low probability of happening.

1.2 Probing the formation of amyloid fibrils

An understanding of all of the mechanisms involved in the formation of amyloid fibrils is necessary if their formation is to be controlled for use as novel materials or in disease prevention. The initial stages of self-assembly are of particular importance in both cases, owing to their apparently low probability of occurring and the possible toxicity of small aggregates. Unfortunately, these are also the hardest part of the process to probe, and so they often have to be inferred from measurements taken later in the self-assembly process, once fibrils have become large enough to be detectable.

Images of mature fibrils are much easier to obtain than those of early-stage aggregates. Electron microscopy and atomic force microscopy reveal apparent differences in fibril length, thickness, and flexibility, depending on the environmental conditions and protein from which the fibrils formed. The images in figure 1.2 were obtained by my collaborator, Ryan Morris using a transmission electron microscope (**TEM**). The image on the left is for bovine insulin, while

(a) Fibrils formed from bovine insulin.



(b) Fibrils formed from ovalbumin.

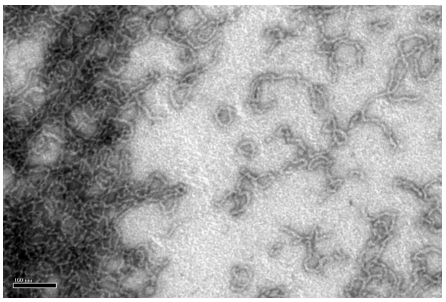


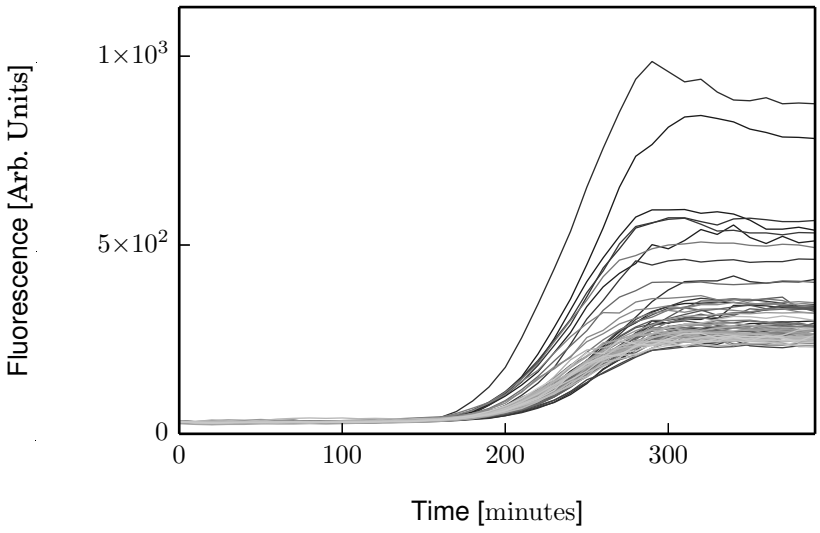
Figure 1.2 Transmission electron micrographs of amyloid fibrils formed from **(a)** bovine insulin, and **(b)** ovalbumin. These samples were taken from experiments after the fluorescence signals had plateaued, by my collaborators Ryan Morris [74] and Jason Kalapothakis. Both images have a 0.1 μm scale bar.

that on the right is for ovalbumin, highlighting the contrasting appearance of fibrils from different proteins. From images such as these, distributions of fibril lengths can be obtained, as well as differences in the apparent fibril structure. For example, under the conditions which produced these images, the persistence length of insulin fibrils is much greater than that of ovalbumin. This information can provide corroborative insight into the mechanisms at work during the fibril's formation. However, taking images of growing fibrils is extremely difficult and has only been achieved a few times using an atomic force microscope [34, 66, 69].

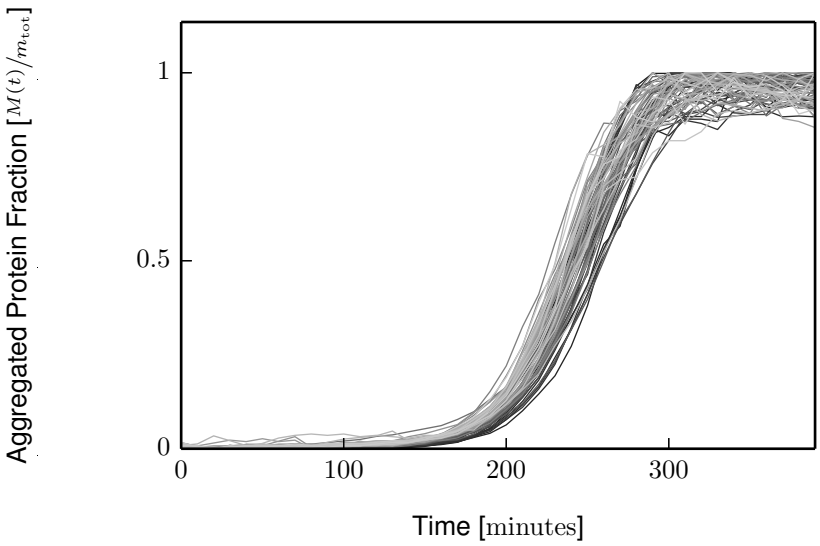
Much simpler techniques exist which can provide time course data on the aggregation process, albeit with little or no spatial resolution. These include light scattering [38, 64], turbidity [38, 90], and fluorescence and absorption spectroscopy [75]. One of the most commonly used techniques employs the dye Thioflavin T (**ThT**) which is known to fluoresce when attached to proteins containing β-sheet structures, such as the cross-β amyloid fibril structure. By measuring the change in fluorescence intensity, a change in the prevalence of β-sheet, and hence amyloid fibril, can be inferred. This technique allows the time course of the self-assembly process, also known as the kinetic curve or growth curve, to be established.

The relationship between the concentration of fibrillar protein and the strength of the ThT fluorescence signal is not generally known and is likely to be specific to the combination of dye concentration, protein, environment, and instrument used

(a) Growth curves of recorded ThT fluorescence values in arbitrary units.



(b) Growth curves of fluorescence values normalised by the maximum in each growth curve.



Replicate bovine insulin experiments

Figure 1.3 Replicate growth curves for the self-assembly of amyloid fibrils from $131\ \mu\mathrm{M}$ bovine insulin. These 68 replicates were obtained simultaneously, and are shown in terms of **(a)** the recorded, and **(b)** the normalised fluorescence values, where the latter indicates the fraction of total protein that has been incorporated into fibrils. The experiments were performed by my collaborator Ryan Morris, using a BMG Labtech Fluostar plate reader, by measuring ThT fluorescence at $10\ \mathrm{minute}$ intervals, with each $100\ \mu\mathrm{l}$ sample incubated at $60^{\circ}\mathrm{C}$ in $25\ \mathrm{mM\ HCl}$ [74].

to measure the fluorescence. Because of this, the easiest method for calibrating the fluorescence resulting from a particular concentration of fibrillar protein is to make use of the long time "end-point" state, where the fluorescence signal has reached a plateau and aggregation appears to have stopped. At this point, the concentration of protein in amyloid fibrils can be determined accurately using other methods, such as measuring the concentration of protein still in solution. These often reveal that all except a negligible fraction of the total protein concentration has been incorporated into fibrils. Figure 1.3b shows examples of growth curves that have been normalised in this way [74], where the maximal (end-point) fluorescence values correspond to almost complete conversion of protein into fibrils; hence, these growth curves are thought to indicate the fraction of the starting protein concentration that has been converted into fibrils, as a function of time.

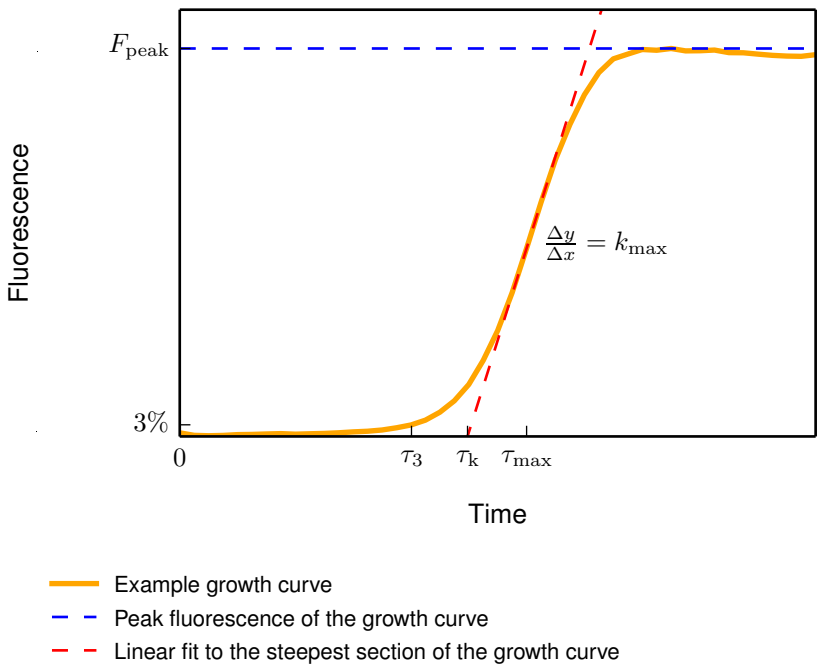


Figure 1.4 Illustration of a typical sigmoidal growth curve. This is annotated with the definitions of the maximum fluorescence value F_{peak}, the lag times τ_{k} and $\tau_{\phi_M=0.03}$, the maximum aggregation rate k_{max}, and the point in time τ_{max} when the maximum aggregation rate is reached.

These kinetic measurements often result in a characteristic sigmoid-shaped growth curve, as illustrated in figure 1.4. The quiescent lag phase at early times may be more or less prominent for different protein systems, and can be

absent in some. Following the lag phase there is a rapid increase in aggregation rate, followed by an inflection point after which aggregation slows and tends to a plateau with a maximum fluorescence value of F_{peak}. The inflection point, between the increasing and decreasing aggregation rates, is termed the point of maximum aggregation and is typically seen to occur at round 50% of the plateau value; in figure 1.4 this point is labelled with the gradient k_{max}, which occurs at the time τ_{max}. The plateau almost always represents complete incorporation of the free monomer into amyloid fibril, and this was indeed the case for bovine insulin under the conditions used to produce the data in figure 1.3.

The length of the lag phase is commonly referred to as the lag time (τ_{lag}) and can be defined in a number of ways; the two most common being to measure the time to reach a threshold fraction of the plateau value, or to linearly extrapolate the slopes at the point of maximum aggregation point down to the time axis. These are labelled as τ_3 (for 3% aggregation) and τ_{k} respectively in figure 1.4.

Several theories have been proposed about the mechanisms which could give rise to such sigmoidal growth curves, and in particular the nature of the lag phase [17, 54, 72]; these will be discussed next.

1.3 Mechanistic models for amyloid fibril polymerisation

Amyloid fibrils, like most linear polymers, typically grow by the addition of individual protein "monomers". However, most proteins need to change conformation into a "misfolded state" (as described in §1.1) before they can be incorporated into a fibril. For most proteins, this does not happen spontaneously on time scales accessible experimentally, without external influences such as an increase in temperature. Even under these conditions however, the self-assembly kinetics of many proteins exhibit a considerable lag before a significant amount of protein becomes incorporated into fibrils. This lag phase can be removed, as illustrated in figure 1.5, by inoculation or "seeding" with mature fibrils.

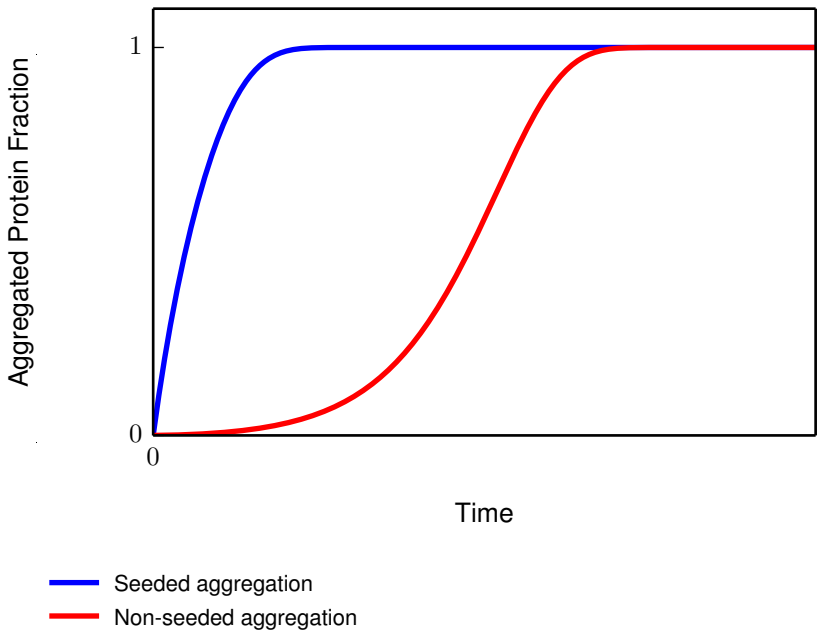


Figure 1.5 A sketch of two growth curves to illustrate the effect of "seeding" amyloid fibril self-assembly with preformed fibrils. The typical sigmoidal growth curve **(red)** displays a lag phase at early times that can be removed by inoculating the sample with preformed fibrils at $t = 0$. This results in a growth curve with a shape similar to an exponential approach **(blue)**.

These factors have lead to the widely accepted view that amyloid fibril formation is a nucleation-dependent polymerisation process, where the *de novo* creation of fibrils occurs via a homogeneous, primary nucleation mechanism (henceforth simply referred to as primary nucleation), and elongation occurs via reversible polymerisation at both fibril ends. The term "primary" is used to distinguish this from other types of nucleation process which appear in more complex models, and also to indicate that (in most cases) it is thought to be responsible for initiating the self-assembly process. As in classical nucleation theory, homogeneous nucleation refers to nucleation in bulk solution, in contrast to heterogeneous nucleation which is catalysed by the presence of a surface or interface as described later in this chapter.

1.3.1 Nucleation-dependent polymerisation

The nucleation-dependent polymerisation model was first introduced by Oosawa and Kasai [82] to describe the assembly of actin and tubulin filaments, and was later improved by Hofrichter [45] to describe the aggregation of sickle cell haemoglobin. The model consists of two processes: the formation of a primary nucleus consisting of n_c monomers, and its subsequent growth by reversible polymerisation.

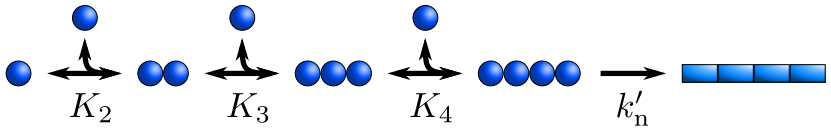


Figure 1.6 A cartoon depicting the formation of a primary nucleus (the shortest stable filament) via a sequence of monomer addition steps (with equilibrium rate constants K_i, where i is the number of monomers involved in the resulting aggregate) and an irreversible conformational change (with rate constant k'_n). Later, this entire sequence will be considered as a single step process with a single rate constant, k_n. This corresponds to reaction 1.1.

In these models, it is assumed that the nucleus forms through a series of reversible polymerisation and conformational rearrangement processes (illustrated by figure 1.6), some or all of which may be energetically or entropically unfavourable. This type of nucleation processes is closely related to classical nucleation theory, originally used to explain crystal formation in supersaturated solutions. It should be noted that the definition of what constitutes the "nucleus" is often ambiguous, as the rate limiting step — prior to becoming a growth competent fibril — may constitute the addition of a final monomer or a conformational rearrangement, and the nucleus size may thus be chosen with respect to either side of this final step. Throughout this thesis — and in almost all other work — it is assumed that the processes which lead to this rate limiting step are comparatively rapid and that their products are in equilibrium with the free monomer population; however, in most cases I will neglect these intermediates when considering nucleation processes. For clarity in later discussion, I choose to define the nucleus as the smallest stable, growth competent species consisting of n_c monomers. This constrains the nucleus size to $n_c \geq 2$ due to the fact that a conformational change in an aggregate of size n_c to produce the nucleus must

be equivalent to the addition of a monomer to an aggregate of size $n_c - 1$, and hence allows the equilibrium constants and the rate constant for the final step to be subsumed into a single nucleation rate constant k_n.

These concepts — the one-step formation of a stable, growth competent nucleus comprising n_c monomers, and equilibrium of all filaments containing fewer than n_c monomers with the free monomer population — can be written as the following two reactions.

$$n_c \cdot m \xrightarrow{k_n m(t)^{n_c}/n_c!} f_{n_c} \tag{1.1}$$

$$f_{i<n_c} \longrightarrow i \cdot m \tag{1.2}$$

In reaction 1.1, n_c monomers are converted into a nucleus with a rate proportional to the free monomer concentration $m(t)$ raised to the power n_c and rate constant k_n as defined above. The term $m(t)^x$ comes from $x - 1$ monomer addition steps involved in going from a single monomer to an aggregate of size x. The order in which these monomers come together is not important and so the factor $n_c!$ is introduced to correct for the number of permutations.

Reaction 1.2 simply states the condition that all aggregates smaller than the nucleus are in equilibrium with the free monomer population and can be considered as part of that population. This will be needed in the event that a filament becomes shorter than the minimum stable length n_c via either depolymerisation or filament breakage, both of which will be considered later.

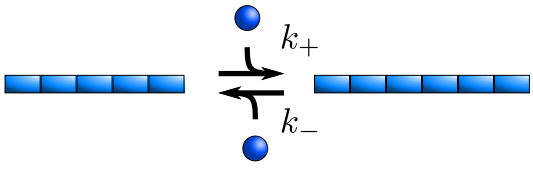


Figure 1.7 A cartoon depicting reversible polymerisation, where elongation by monomer addition occurs with rate constant k_+, and depolymerisation of monomer occurs with rate constant k_-. These correspond to reactions 1.3 and 1.4, respectively.

Reversible polymerisation of a filament f_i containing i monomers can be modelled by the following two reactions:

$$f_i + m \xrightarrow{2k_+ m(t)} f_{i+1} \tag{1.3}$$

$$f_i \xrightarrow{2k_-} f_{i-1} + m \tag{1.4}$$

In reaction 1.3, a filament grows longer by the addition of one monomer m, with a rate proportional to the free monomer concentration $m(t)$ and rate constant k_+. In reaction 1.4, one monomer is removed from a filament of length i, with a rate proportional to the rate constant k_-. The factor of 2 in both reactions is to account for the fact that fibrils can grow and shrink from both ends. These are illustrated together in figure 1.7.

A set of time-dependent differential equations for the concentrations of fibrils of length f_i can be constructed from reactions 1.1, 1.2, 1.3, and 1.4. Oosawa and Kasai [82] showed that this system of equations could be solved by assuming that the formation of nuclei was limited only by the rate of the final conformation step shown in figure 1.6 and that this was a slow process such that the monomer population would be much larger than the population of nuclei at early times; i.e. $m(t) = m_{\text{tot}} - M(t) \simeq m_{\text{tot}}$ where $M(t)$ is the concentration of fibrillar protein at time t. These assumptions guarantee that depolymerisation can be neglected, making it possible to find a closed-form analytic description of the time evolution of the fibrillar protein concentration $M(t) = \sum_{\forall i \geq n_c} i f_i(t)$ [82, 85]. Their result is commonly written as

$$M(t) = m_{\text{tot}} \left[1 - \operatorname{sech}^{2/n_c} \left(\sqrt{\frac{n_c}{2}} \delta t \right) \right] \tag{1.5}$$

where $\delta = \sqrt{2k_+ k_n m_{\text{tot}}^{n_c} / n_c!}$. This equation produces sigmoid-like growth curves with a time scale determined by the parameter δ^{-1}. From this parameter the scaling behaviour of the lag time with total protein concentration can be shown to be $\tau_{\text{lag}} \sim m_{\text{tot}}^{-\gamma}$ where $\gamma = {}^{n_c}/_2 \geq 1$ because $n_c \geq 2$ according the to definition chosen earlier.

Thus, the nucleation-dependent polymerisation model makes a strong prediction about the way in which the lag phase will scale with total protein concentration. However, in contrast to this prediction, a number of amyloid fibril forming

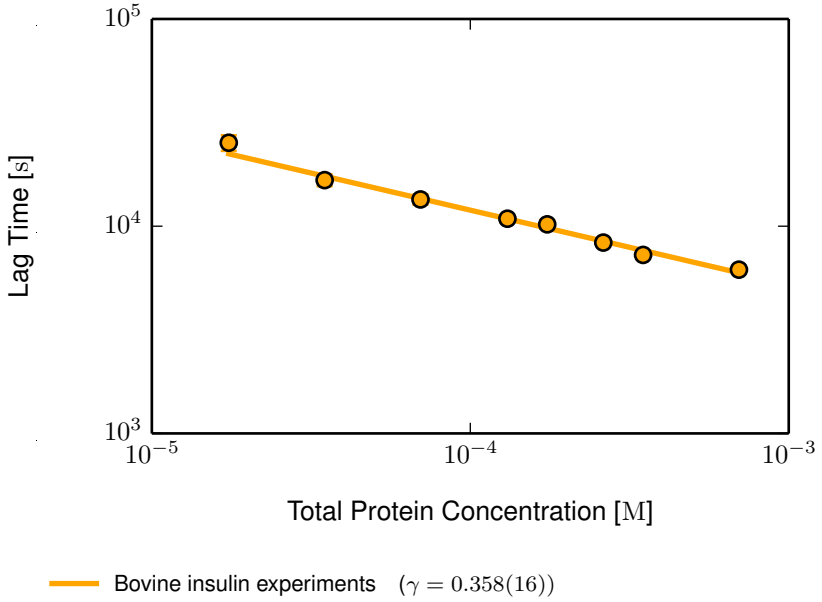


Figure 1.8 The scaling behaviour of τ_3 with m_{tot} in bovine insulin experiments; each point represents the mean lag-time of between 140 and 200 replicate experiments at a particular protein concentration with error bars (smaller than the symbol size) indicating the standard deviation. The line is the best fit of the power law $\tau_{\mathrm{lag}} \propto m_{\mathrm{tot}}^{-\gamma}$ to these data; the value $\gamma < 1$ indicates that bovine insulin fibril formation can be better described by the fragmentation model (§1.3.4), than either the primary, or heterogeneous nucleation models. The lag times (extracted from growth curves such as those shown in figure 1.3) and value of γ were both determined using to the methods described in §2.3.2.

proteins have been reported to give a value of $\gamma < 1$ [15, 18, 55, 58]; for example, bovine insulin as shown in figure 1.8. Moreover, for some of those protein systems where $\gamma < 1$, the shape of the growth profile at early times is found to be steeper than predicted by this model [26]; i.e. the time evolution of the fibrillar protein concentration $M(t)$ is better described by an exponential than by the parabolic approximation ($M(t) \sim t^2$) of equation 1.5. These and other observations have led to many additions and modifications to this nucleation-dependent polymerisation model, one of which I will now describe.

1.3.2 Autocatalytic polymerisation

Autocatalytic polymerisation involves a feedback mechanism by which existing polymer stimulates the formation of new polymer. In this context, the

concentration of polymerised material acts like a catalyst, enhancing the rate of formation of new growth-competent ends beyond that normally produced by primary nucleation; for example, the surface of an existing polymer may act as a template to which free monomers can bind and form a new fibril (shown below in figure 1.9). By creating additional growth-competent ends, the rate at which free monomers are added to polymer ends, is increased. When additional growth-competent ends form at a rate proportional to the concentration of polymerised material, then a higher rate of polymerisation will lead to a higher rate of end formation, and *vice versa*.

A number of autocatalytic mechanisms have been proposed to explain aspects of amyloid fibril formation which depart from the nucleation-dependent polymerisation model [25, 106]. The two which are most relevant to this thesis are heterogeneous secondary nucleation and filament fragmentation [16].

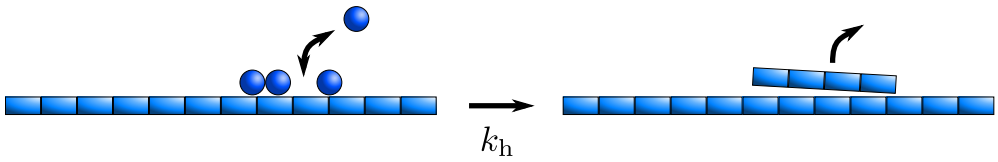


Figure 1.9 A cartoon depicting the heterogeneous nucleation of a new filament. Here an existing filament acts as a template to which free monomers can reversibly bind until n_s monomers are in close enough proximity and in the appropriate conformation to form a stable fibril. This involves similar steps to those during the formation of a primary nucleus in figure 1.6, and this sequence is also treated as a single step with a single rate constant, k_h, as given by reaction 1.11.

In the 1980s, Ferrone et al. [25] noted that sickle haemoglobin aggregation progressed more rapidly than predicted by the nucleation-dependent polymerisation model, and proposed that heterogeneous, secondary nucleation of new fibrils was occurring on the surface of mature fibrils: this is illustrated in figure 1.9 and will be hereafter referred to simply as heterogeneous nucleation. They proposed a modification to the nucleation-dependent polymerisation model which accounted for autocatalytic formation of new, growth-competent filaments, at a rate proportional to the current concentration of fibrillar protein. This mechanism succeeded in reproducing a lag phase and exponential early growth profile, but only to first order in their analytic solution [26], noting that higher order iteration and exact numerical solutions underestimated the contribution from secondary nucleation and hence recovered the parabolic approximation of equation 1.5.

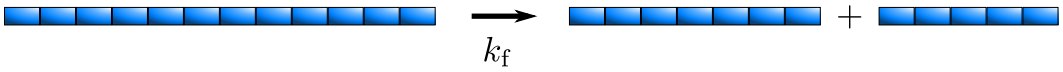


Figure 1.10 A cartoon depicting filament fragmentation where one of the junctions between the monomers of a filament, breaks, with the rate constant k_f. So long as both fragments are of a stable length, this results in an additional, growth competent filament, corresponding to reaction 1.10. If either, or both fragments are unstably short, then they are expected to disintegrate into free monomer in accordance with reaction 1.2.

At around the same time Wegner and Savko [106] developed an alternative autocatalytic model, to describe the self-assembly of actin filaments. Again based on the nucleated-polymerisation model, they incorporated filament fragmentation as a secondary process: this is illustrated in figure 1.10. Again, new filament ends are created at a rate proportional to the current concentration of fibrillar protein, but unlike the heterogeneous nucleation mechanism of Ferrone et al. [26], the rate of fragmentation does not depend on the free monomer population since a filament is capable of breaking and forming two additional growth-competent ends without consuming free monomer. In 2009, Knowles et al. [55] published an analytic solution to a similar fragmentation model, describing the change in fibrillar protein concentration over time, in the mean-field limit. This solution reproduces the sigmoidal growth curves and crucially predicts that their lag times will scale with total protein concentration according to a power law with exponent $\gamma = {}^1\!/_2$.

In a later series of publications, Cohen et al. [15] showed that both fragmentation and heterogeneous nucleation could be treated with similar analytic techniques and thus result in similar, mean-field expressions for the time evolution of the fibrillar protein concentration [12, 13, 15]. The full autocatalytic model they presented involves the set of reactions in table 1, where f_i represents a filament containing i monomers.

Reactions 1.6, 1.7, 1.8, and 1.9 in table 1 are clearly identical to those of the nucleation-dependent polymerisation model in §1.3.1. The two additions are reaction 1.10, which describes the fragmentation of a filament with rate constant k_f, and reaction 1.11 which describes the heterogeneous nucleation of a filament with rate constant k_h.

$$n_{\mathrm{c}} \cdot m \xrightarrow{k_{\mathrm{n}} m(t)^{n_{\mathrm{c}}} / n_{\mathrm{c}}!} f_{n_{\mathrm{c}}} \tag{1.6}$$

$$f_{i<n_{\mathrm{c}}} \longrightarrow i \cdot m \tag{1.7}$$

$$f_i + m \xrightarrow{2k_+ m(t)} f_{i+1} \tag{1.8}$$

$$f_i \xrightarrow{2k_-} f_{i-1} + m \tag{1.9}$$

$$f_i \xrightarrow{(i-1)k_{\mathrm{f}}} f_j + f_{i-j} \tag{1.10}$$

$$f_i + n_{\mathrm{s}} \cdot m \xrightarrow{i k_{\mathrm{h}} m(t)^{n_{\mathrm{s}}} / n_{\mathrm{s}}!} f_i + f_{n_{\mathrm{s}}} \tag{1.11}$$

Table 1 The reaction set for a model for autocatalytic amyloid fibril self-assembly: primary nucleation (1.6), disintegration of filaments of unstable length (1.7), polymerisation (1.8), depolymerisation (1.9), fragmentation (1.10), heterogeneous nucleation (1.11). Where they appear, n_{c} is the critical nucleus size and n_{s} is the heterogeneous nucleus size, and k_{n}, k_+, k_-, k_{f}, and k_{h} are the rate constants for primary nucleation, elongation, depolymerisation, fragmentation, and heterogeneous nucleation, respectively.

Heterogeneous nucleation forms a new filament $f_{n_{\mathrm{s}}}$, when n_{s} monomers come together on the surface of an existing filament. In the same way as primary nucleation, this can be approximated by a single step with rate proportional to $m(t)^{n_{\mathrm{s}}}$, and the filament populations $f_{i<n_{\mathrm{s}}}$ are assumed to also be in rapid equilibrium with the free monomer population. Similarly, the factor $n_{\mathrm{s}}!$ in reaction 1.11 is included to correct for the indistinguishability of the monomers in the resulting filament.

In the fragmentation reaction, a filament f_i is effectively destroyed in order to create two new filaments, f_j and f_{i-j}. Unlike heterogeneous nucleation, this does not appear catalytic because the filament which forms the new growth-competent ends is effectively destroyed by the process. However, as I stated earlier, it is the total concentration of aggregated protein that acts as the catalyst, and this remains unchanged.

The propensities of both fragmentation and heterogeneous-nucleation mechanisms are linearly proportional to the length of the filament which aided in the formation of a few filament; the assumption being that longer filaments

have a higher probability of breaking, or equivalently a larger surface on which heterogeneous nucleation can occur. The difference in the length dependence of the two mechanisms comes from the fact that fragmentation can occur at $i-1$ places along a fibril of length i (i.e. between the monomers), while heterogeneous nucleation can occur at i places (i.e. on the monomers). Alternatives to this linear scaling of the mechanisms with filament length have been proposed and in some cases solved [44, 47]. However, they are considerably more complex and do not provide a significant improvement over the original, linear assumption in terms of fitting to experimental data.

It is also worth pointing out at this juncture that fragmentation can cause the loss of a single monomer from the end of a growth-competent filament. This is not accounted for in the heterogeneous nucleation model without also considering depolymerisation of monomer from all stable filaments (reation 1.9). If either process results in a filaments containing fewer than n_c monomers, these disintegrate in accordance with reaction 1.7.

1.3.3 A mean-field analytic solution to the autocatalytic polymerisation model of amyloid fibril self-assembly

The autocatalytic reaction scheme presented in table 1, can be summarised as: primary nucleation of a filament of size n_c; equilibrium of all aggregates containing fewer than n_c monomers with the free monomer population; growth of filaments by polymerisation of free monomer to their ends; shrinkage of filaments by depolymerisation of monomer from their ends; autocatalytic generation of new, growth-competent filament ends by fragmentation of existing filaments; and autocatalytic creation of new filaments containing n_s monomers by heterogeneous nucleation on the surface of existing filaments, dependent on the free monomer concentration.

Starting from this set of reactions, Cohen et al. [15] constructed a set of differential equations for the evolution of the concentration of fibrils of length i, $\{f_i\}$ [12, 13, 15]. By summing this set over all lengths, differential equations for the first two moments of the filament population, $N(t)$ and $M(t)$, were constructed. These describe the change in the number concentration of growth-competent filaments (or half the number of growth-competent ends), and the

concentration of protein in these fibrils (the concentration of monomers within fibrils), respectively:

$$\frac{dN(t)}{dt} = k_{\mathrm{f}}\big[M(t) - (2n_{\mathrm{c}} - 1)N(t)\big] + k_{\mathrm{h}}M(t)m(t)^{n_{\mathrm{s}}} + \frac{k_{\mathrm{n}}m(t)^{n_{\mathrm{c}}}}{n_{\mathrm{c}}!} \tag{1.12}$$

$$\frac{dM(t)}{dt} = 2\big[m(t)k_{+} - k_{-} - n_{\mathrm{c}}(n_{\mathrm{c}} - 1)k_{\mathrm{f}}\big]N(t) + n_{\mathrm{s}}k_{\mathrm{h}}m(t)^{n_{\mathrm{s}}} + \frac{n_{\mathrm{c}}k_{\mathrm{n}}m(t)^{n_{\mathrm{c}}}}{n_{\mathrm{c}}!} \tag{1.13}$$

where $m(t)$ is the time dependent free monomer concentration and all other symbols are as defined for the reactions in table 1. The first term in equation 1.13 describes the change of fibrillar protein concentration, which is increased by monomer addition (elongation), and reduced by depolymerisation and the formation of fragments containing fewer than n_{c} monomers. Similarly, the second term in equation 1.13 describes the increase in fibrillar protein concentration when a new fibril is formed by heterogeneous nucleation. Thus the first two terms in equation 1.12 describe the increase in fibril concentration via fragmentation and heterogeneous nucleation, respectively; and the last term in both equations accounts for the increase in the concentration of fibrils and fibrillar protein due to primary nucleation.

The pair of equations 1.12 and 1.13 were then solved iteratively, for each of the autocatalytic mechanisms independently (i.e. with either of k_{f} or k_{h} set to zero). This was done using the technique of fixed point analysis, by neglecting depolymerisation, taking the early-time approximation $m(t) = m_{\mathrm{tot}} - M(t) \simeq m_{\mathrm{tot}}$, and assuming monomer consumption by primary nucleation terms $\mathcal{O}(k_{\mathrm{n}})$ could be neglected. Good agreement with numerical solutions of equations 1.12 and 1.13 was reported after one iteration in the case of $M(t)$, and two iterations in the case of $N(t)$. This approach was first used by Knowles et al. [55] for the fragmentation model alone, but is presented more generally for both autocatalytic models in later work by Cohen et al. [12]. The additional non-linearity introduced by the heterogeneous nucleation mechanism means that further iterations of the above technique are required in order to produce the same quality of agreement with numerical solutions as was seen for the fragmentation model [12]. However, these more exact solutions provide little extra insight into the behaviour of the system and will not be considered in this thesis.

The form of $M(t)$ which is obtained by the above method can be expressed generally by equation 1.14, without reference to either k_f or k_h (as is explained below).

$$M(t) = m_\mathrm{tot}\left(1 - \exp\left(-C_+ e^{\kappa t} + C_- e^{-\kappa t} + \frac{\delta^2}{\kappa^2}\right)\right) \tag{1.14}$$

$$C_\pm = \frac{k_+ N(0)}{\kappa} \pm \frac{M(0)}{2m(0)} \pm \frac{\delta^2}{2\kappa^2} \tag{1.15}$$

$$\delta = \sqrt{2k_+ k_\mathrm{n} m_\mathrm{tot}^{n_\mathrm{c}}/n_\mathrm{c}!} \tag{1.16}$$

The parameter δ, which also emerged in the nucleation-polymerisation model in §1.3.1, is the characteristic rate for primary nucleation driven polymerisation. This determines the way that the growth curve scales with total protein concentration when primary nucleation is dominant over autocatalysis; i.e. recovering the prediction of the nucleation-dependent polymerisation model.

The parameter κ similarly describes the characteristic rate for autocatalytic polymerisation, and its form depends on the specific mechanisms:
For fragmentation (where $k_\mathrm{f} > 0$ and $k_\mathrm{h} = 0$)

$$\kappa = \sqrt{2k_+ k_\mathrm{f} m_\mathrm{tot}} \tag{1.17}$$

and for heterogeneous nucleation (where $k_\mathrm{h} > 0$ and $k_\mathrm{f} = 0$)

$$\kappa = \sqrt{2k_+ k_\mathrm{h} m_\mathrm{tot}^{n_\mathrm{s}+1}/n_\mathrm{s}!} \tag{1.18}$$

Both forms of κ contain the propensity $2k_+ m_\mathrm{tot}$ [(2)] from reaction 1.8, along with either the propensity for fragmentation k_f (from reaction 1.10), or the propensity for heterogeneous nucleation $k_\mathrm{h} m_\mathrm{tot}^{n_\mathrm{s}}/n_\mathrm{s}!$ (from reaction 1.11).

The rate at which self-assembly proceeds is thus determined by which of the parameters δ or κ is largest. In §5, the competition between these rates will be explored in more detail, but for most of this thesis it can be assumed that autocatalysis, and hence κ, is dominant.

(2) Here, I have used the early-time approximation $m(t) \approx m_\mathrm{tot}$.

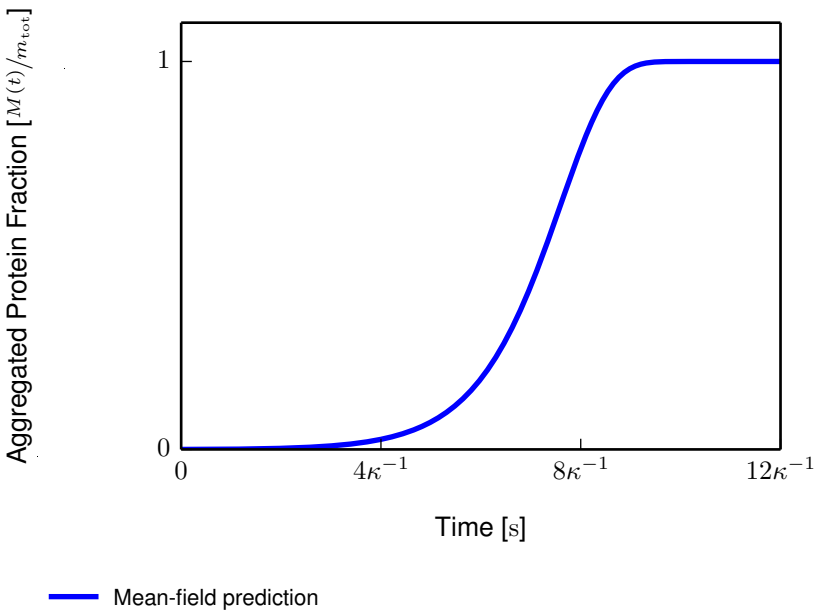


Figure 1.11 A growth curve produced by equation 1.14, plotted in terms of the characteristic time scale κ^{-1}, where on the same time axis $\delta^{-1} \approx 32\kappa^{-1}$ cannot be seen.

Equation 1.14 produces sigmoid-like growth curves (figure 1.11), with initially exponential growth and a more rapid change in aggregation rate than those of the nucleation-dependent polymerisation model (equation 1.5).

1.3.4 Predictions of the mean-field analytic description of autocatalytic self-assembly

In addition to the above expression for the growth curve (equation 1.14), this mean-field description also provides a way to access analytic expressions for particular aspects of these growth curves, such as the maximum aggregation rate and lag time, and also other aspects of the aggregation kinetics, such as the mean filament length.

Maximum aggregation rate

The maximum aggregation rate can be obtained from equation 1.14 by evaluating $\frac{dM(t)}{dt}$ at the time of the inflection point $t_{\max} = \frac{\log\left(\frac{1}{C_+}\right)}{\kappa}$, which is obtained

from the solution to $\frac{d^2M(t)}{dt^2} = 0$. This gives,

$$k_{\max} = M(\infty)\kappa e^{-1} \tag{1.19}$$

which is dependent on the characteristic rate parameter κ, and independent of the initial conditions $M(0)$ and $N(0)$.

This predicts that the maximum aggregation rate scales with total protein concentration according to $k_{\max} \sim m_{\text{tot}}^{1+\gamma}$, where $\gamma = {}^{(n_s+1)}/_2$ with secondary nucleus size n_s (where $n_s = 0$ for fragmentation). However, more complex scaling behaviour has been reported experimentally [74], which will be investigated in §6.

Lag-time expressions

Two expressions for the mean lag-time (illustrated in figure 1.4) can be derived from equation 1.14. These are the time τ_{ϕ_M} to reach a threshold fraction of the total protein concentration, and the time τ_k which is obtained by extrapolation of the maximum aggregation rate.

The time τ_{ϕ_M} at which a threshold fraction of aggregate $M_\phi = M(t_{\text{threshold}}) \approx \phi_M\, m_{\text{tot}}$ is reached can be found from 1.14 to be:

$$\tau_{\phi_M} = \frac{1}{\kappa} \ln\left(\frac{D - \Phi + \sqrt{(\Phi - D)^2 + 4C_+C_-}}{2C_+}\right) \tag{1.20}$$

where $D = \frac{\delta^2}{\kappa^2}$ and $\Phi = \ln(1 - \phi_M)$. This can be simplified for early self-assembly, where $\phi_M \ll 1$ (i.e. $\Phi \to 0$), and when autocatalysis dominates the kinetics, where $\delta \ll \kappa$ (i.e. $D \approx 0$), to give

$$\tau_{\phi_M} = \frac{\log\left(\frac{\phi_M}{C_+}\right)}{\kappa} \tag{1.21}$$

The lag time found by linear extrapolation of the maximum aggregation rate can be obtained analytically from the definitions of $k_{\max}$, $t_{\max}$ and $M(t_{\max})$:

$$\tau_k = t_{\max} + \frac{1 - e}{\kappa} = \frac{\log\left(\frac{1}{C_+}\right) + 1 - e}{\kappa} \tag{1.22}$$

Both lag time expressions 1.21 and 1.22 are almost identical and are both inversely proportional to κ. This makes sense since κ^{-1} and δ^{-1} are the only time scales involved. When $M(0)$, $N(0)$, and δ are all small, the dominant contribution to the protein concentration dependence of the lag time will come from κ. Hence, the autocatalytic model predicts that the lag time will scale with total protein concentration according to $\tau_{\text{lag}} \sim m_{\text{tot}}^{-\gamma}$, but in this case $\gamma = {}^{(n_s+1)}/_2$ depends on the autocatalytic mechanism involved: in the case of heterogeneous nucleation where $n_s > 0$, $\gamma \geq 1$; whereas for fragmentation where $n_s = 0$, $\gamma = \frac{1}{2}$.

This fragmentation model was the first to reproduce a lag-time scaling exponent $\gamma < 1$, which is observed experimentally in a number of protein systems [18, 27, 53, 55, 94, 117]. However, in order to obtain this scaling, fragmentation-dominated polymerisation, rather than primary nucleation, must entirely determine the lag phase as will be shown in §5.

Equations 1.20 and 1.22 only provide mean lag-times and do not address the origin of the variability often observed experimentally in the growth curves. Naively, this might be thought to reflect underlying fluctuations in the individual mechanisms of primary nucleation and autocatalytic polymerisation. However, as will be shown in §4.2, contributions to the lag-time variability from these mechanisms become vanishingly small under the conditions typically used in laboratory experiments.

Fibril length distribution

For the nucleation-dependent polymerisation model, Oosawa and Asakura [81] showed that a differently shaped distributions of filament lengths could be expected to emerge at different time points in the kinetics. Once the filament population reaches equilibrium with the free monomer population, they predict that a Poisson distribution of filament lengths will emerge and persists while the equilibrium is maintained. After the monomer population has become depleted, this distribution is expected to evolve into an exponential distribution described by the parameter combination $\frac{m(\infty)k_+}{k_-}$, at long times. This relaxation is expected to occur via a diffusion-like process where depolymerisation of long fibrils — which became abundant during the growth process — release material which

contributes to the growth of shorter fibrils, and gradually shifts the population towards shorter filament lengths.

For heterogeneous nucleation with reversible polymerisation, Cohen et al. [13] predicts almost identical behaviour, recovering the same analytic result as Oosawa and Asakura [81] when $k_{\mathrm{h}} = 0$. From this, they postulated that the qualitative shape [(3)] of these length distributions is an inevitable product of a nucleation processes which produces filaments of a uniform length, and which satisfies detailed balance in the exchange processes between filaments of different lengths.

The fragmentation model, by contrast, permits the formation of filaments of any length and does not satisfy detailed balance without filament re-association. This leads to very different prediction for the length distributions. In the long time limit, Cohen et al. [13] found that the length distribution for the fragmentation model could be described by a "biased Gaussian" (skew normal) determined almost entirely by the parameter n_{c} [13] [(4)]. Earlier in the self-assembly process (when $m(t) \approx m_{\mathrm{tot}}$), a broader, skew normal distribution is predicted to exist, controlled by the parameter combination $\sqrt{\frac{2m_{\mathrm{tot}}k_+}{k_{\mathrm{f}}}} = \frac{\kappa}{k_{\mathrm{f}}}$.

The shape and evolution of these distributions can be understood through competition between the mechanisms of elongation and either fragmentation or heterogeneous nucleation [13]. Under the action of fragmentation, filaments will tend to be broken into their smallest stable size, n_{c}. Fragments smaller than n_{c} will enter equilibrium with the free monomer population and are re-polymerised onto existing filaments. Similarly, depolymerisation will cause existing filaments to become shorter until they, too, are no longer stable, leading to the same disintegration and re-polymerisation behaviour. Polymerisation will act to oppose both of these so long as sufficient free monomer is available.

(3) The length distribution for heterogeneous nucleation evolves over time, from one which has a peak into an exponential.

(4) The peak position tends to n_{c} as $n_{\mathrm{c}} \to 2$ such that when $n_{\mathrm{c}} = 2$ the distribution can be approximated as an exponential.

The peaked length distributions of the intermediate state in both models agree much better with many of those seen experimentally, than do the exponential distributions that are predicted for the long-time steady state [1, 3, 9, 46, 84, 97]. This suggests that in typical experiments, the length distribution is unable to relax, either through fragmentation or depolymerisation, on the time scales of typical experiments, even though these are much longer than predicted by the model. It has been suggested that interactions between filaments could create a stabilising affect, reducing or eliminating fragmentation [74] and retarding the relaxation process. The impact of disabling fragmentation during aggregation, will be explored in §6.

The length distribution also provides a constraint on the interpretation of κ, allowing extraction of the individual rate constants. The mean filament length is defined for all times as

$$\overline{L}(t) = \frac{M(t)}{N(t)} \tag{1.23}$$

regardless of the model involved. To obtain the filament length distribution during the aggregation process, Cohen et al. [13] made the approximation that $m(t) = m_{\text{tot}}$, resulting in exponential growth expressions for both $M(t)$ and $N(t)$. Curiously, the mean filament length defined by equation 1.23 when using these approximations, turns out to be a constant at later times, given by

$$\overline{L}(t) \simeq \frac{2k_+ m_{\text{tot}}}{\kappa} \tanh\left(\frac{\kappa t}{2}\right) \tag{1.24}$$

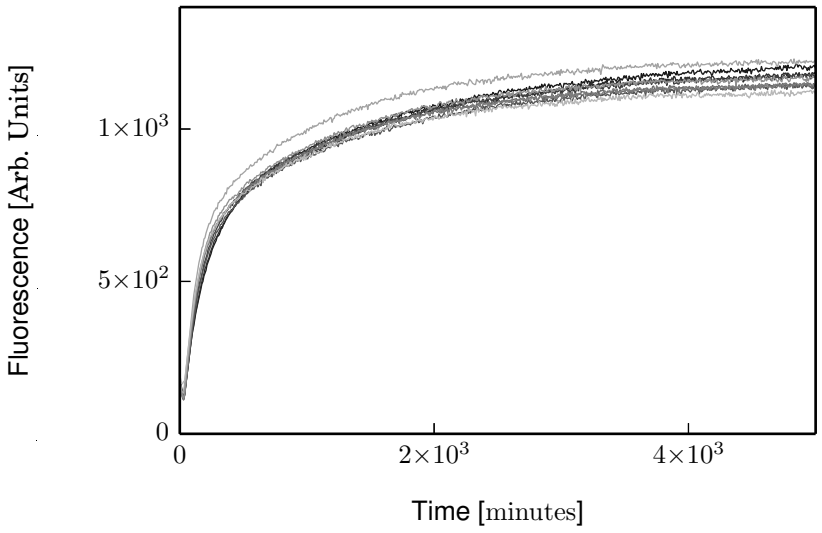
$$\lim_{t\to\infty} \overline{L}(t) = \frac{2k_+ m_{\text{tot}}}{\kappa} \tag{1.25}$$

For experimental data where κ is found to scale with the total protein concentration as predicted by the fragmentation model (equation 1.19), equation 1.25 allows k_+ or k_{f} to be determined from the mean filament length $\overline{L}$.

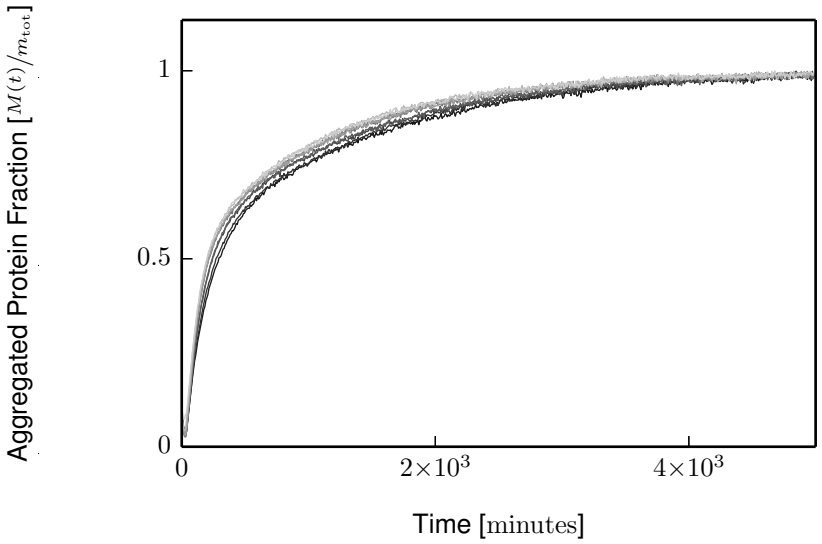
1.3.5 Fibril end-joining and loop formation

A number of amyloid fibril forming proteins display non-sigmoidal kinetics that do not fit with the nucleation-dependent polymerisation or autocatalytic polymerisation models presented above [29, 30, 35, 41, 79, 80, 91, 92, 113, 114]. Many have little or no obvious lag phase, with the maximum aggregation rate

(a) Growth curves of recorded ThT fluorescence values in arbitrary units.



(b) Growth curves of fluorescence values normalised by the maximum in each growth curve.



—— Replicate ovalbumin experiments

Figure 1.12 Replicate growth curves for the self-assembly of amyloid fibrils from $19\ \mu\mathrm{M}$ ovalbumin. Unlike the growth curves of bovine insulin, these do not display a lag phase. These 9 replicates are shown in terms of **(a)** the recorded, and **(b)** the normalised fluorescence values, where the latter indicates the fraction of total protein that has been incorporated into fibrils. The experiments were performed by my collaborators Ryan Morris and Jason Kalapothakis, using a BMG Labtech Fluostar plate reader, by measuring ThT fluorescence at $10\ \mathrm{minute}$ intervals, with each $100\ \mu\mathrm{l}$ sample incubated at $60°\mathrm{C}$ in $10\ \mathrm{mM}$ ammonium acetate and $10\ \mathrm{mM}$ DTT buffer.

occurring close to, or at, $t = 0$. Typically, the lack of a lag phase and rapid early growth are associated with the introduction of pre-formed fibrils, or seeds, at the start [14, 16], but for these proteins this is unnecessary. Example kinetics of this type are shown in figure 1.12.

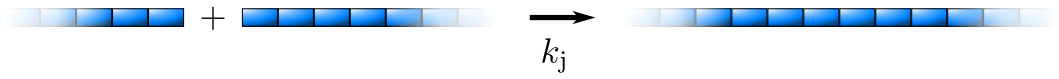


Figure 1.13 A cartoon depicting filament end-joining, with rate constant k_j. The two filament ends may belong to the same filament only if it is flexible enough for the ends to meet; otherwise the two ends will belong to separate filaments.

Filament end-to-end association has been proposed as an alternative mechanism to explain such kinetics [42, 71, 72, 103, 104, 114]. Variously referred to as coagulation, condensation or downhill polymerisation, this mechanism represents an alternative growth pathway to self-assembly [(5)]; it can also be regarded as the reverse of filament fragmentation as illustrated in figure 1.13. In a similar way to polymerisation, I will assume that filament elongation can occur by the addition of another filament to one of its growth-competent ends. This can be summarised by the reaction

$$f_i + f_j \xrightarrow{k_\mathrm{j}} f_{i+j} \tag{1.26}$$

with rate constant k_j. This mechanism should allow more rapid elongation of an individual filament than monomer addition alone. However, it will actually decrease the overall aggregation rate, since it does not change the concentration of fibrillar protein but does decrease the number of growth-competent ends [44].

The corollary to allowing any two filaments to join in this manner is that an individual filament may join to itself, forming a closed loop. Such loops would no longer be growth competent, but would still permit surface processes such as secondary nucleation. Loops have been reported in AFM and electron microscopy experiments alongside non-looped fibrils for a number of protein systems [19, 41, 42, 101, 114], and in many of these cases growth curves also exhibit little or no lag phase and rapid initial growth. Undoubtedly, these loops must form in a length dependent way [42, 57]; Hatters et al. [42] succeeded in modelling the

(5) An alternative to simple monomer addition.

loop frequency observed experimetnally for Apolipoprotein C-II (**apoC-II**) using a worm-like chain closure model [42, 50].

These processes constitute the reverse of filament formation, since they deplete the number of filament ends available for polymerisation by either monomer or filament addition. It is not completely clear what affect this will have on the kinetics, particularly if fragmentation is also involved, but simulations incorporating filament end-joining have shown some success in explaining the distribution of loops and the shape of growth curves produced by apoC-II [42, 114]. While these indicate that end-joining is likely to be the correct explanation, these simulations fall short of providing a detailed picture of the kinetics.

In §7, I present analytic solutions to models incorporating filament end-joining, which allows the rate constants to be estimated from experimental data in much the same way as can be done for the autocatalysis model presented earlier. These also provide predictions for the effects of reversible filament end-joining on the kinetics.

1.4 Going beyond mean-field models

Considerable effort has gone into understanding the mechanisms involved in the formation of amyloid fibrils. Much of this work has focused on the kinetics at early times, which involve short time- and length-scales rendering them difficult to probe experimentally. The early nucleation-dependent model of Oosawa and Asakura [81], and later autocatalytic models of Cohen et al. [17], make strong predictions about how the events occurring at early times are expected to influence observable scaling behaviour later on in the aggregation kinetics.

While these analytic solutions have been very widely used to extract reaction rates and other parameters from experimental data, they only provide a mean-field picture of the kinetics. Substantial variability is commonly observed in replicate experimental growth curves which a mean-field approach cannot explain. This variability is often assumed to reflect inherent fluctuations in the mechanisms involved, and in particular primary nucleation. To verify this assumption, an alternate, stochastic implementation of these models is required.

Kinetic Monte Carlo simulations provide a way to model chemical reaction kinetics, taking into account fluctuations arising from the discreteness of individual molecules and the order of reaction events. In §3, I will present the results of kinetic Monte Carlo simulations for the autocatalytic polymerisation models of Cohen et al. [15], to which I compare the mean-field analytic solutions. I will then explore the origins of the variability in lag time of these and experimental growth curves in §4, by considering the first passage probability distributions for the processes of primary nucleation, polymerisation and autocatalysis.

The scaling behaviour of the lag time with total protein concentration obtained from the mean-field analytic solutions predicts that fragmentation dominated kinetics will scale more weakly than any nucleation dominated kinetics; indeed, it provides an exact value of the exponent, $\gamma = {}^1\!/_2$ for fragmentation dominated kinetics. However, that analysis does not consider individual events. Consequently, self-assembly may proceed via a fragmentation dominated mechanism, but not before the occurrence of other processes with waiting times drawn from a much wider distribution. Using the kinetic Monte Carlo simulation framework developed in §3, along with analytic techniques, I will demonstrate the effect of primary nucleation on this lag-time scaling behaviour, in §5.

The filament length distribution of the fragmentation model changes quite dramatically from a broad peaked distribution throughout most of the kinetics, to a very narrow, almost exponential distribution in the long time limit. The former, peaked distribution has been reported a number of times for systems exhibiting lag-time scaling behaviour otherwise commensurate with the fragmentation model. This suggests that the relaxation of the length distribution is being arrested while still in the peaked form. This will be investigated in §6, where, in order to explain the observed length distribution and transition in the maximum aggregation rate scaling with total protein concentration, I implement a mechanism whereby fragmentation is arrested according to the fraction of total protein currently incorporated into filaments.

Finally, looped fibrils have been reported experimentally, often in conjunction with non-sigmoidal growth curves. The presence of loops indicates that filaments are able to grow by end-joining, and in §7 I will present analytic solutions for models involving filament end-joining with and without fragmentation.

Hence, using kinetic Monte Carlo simulations, and analytic techniques, I will demonstrate how the adjustment of various environmental parameters could be expected to substantially alter the mechanisms which dominate amyloid fibril self-assembly.

Chapter 2

Methods

2.1 Kinetic Monte Carlo Simulation

When choosing which simulation technique to use when modelling a given system (in this case a set of chemical reactions), the trade-off between computational speed, resolution of detail, and comparison to the real system need to be considered in the context of the questions being asked. For example, it would be wasteful to perform a molecular dynamics simulation if the system could be modelled as a well mixed solution, because the extra calculations involved do not contribute significantly to the results.

Kinetic Monte Carlo (Gillespie algorithm) simulations are a useful technique for probing systems of chemical reactions where kinetic phenomena (as opposed to steady state) and the discreteness of individual molecules and events are of interest, but spatial resolution is unimportant. All simulations presented in this thesis were performed using this technique [31, 32].

The Gillespie algorithm assumes a well mixed, homogeneous, isotropic, mixture of active components in an inert solvent (which is ignored). A set of reactions of the form $A + B \xrightarrow{\mu} C$ describe how the components A and B interact to produce C at a rate μ. The propensity of each reaction occurring [1] can be calculated from the abundance of reactants it involves and its rate constant (which is a parameter of the model). For the above example the propensity is given by $\mu X_a X_b$, where μ is the rate constant (with units s^{-1}), and X_x denotes the number of particles of species x. In the case where two molecules of the same species react, this would become $\mu X_a(X_a - 1)$, since the molecule cannot react with itself in this two-body manner.

In this algorithm, each species has an integer number of particles. If needed elsewhere, this can be converted to a concentration $[a]$, using $X_a = [a]V$, where V is a chosen system size [2]. It is important to note that some rate constants of some types of reaction (bimolecular and higher order) will also depend on volume; for example, the rate constant μ will change according to $\mu = \frac{k}{V^{n-1}}$, where n is the number of molecules involved in the reaction [3], and k is the more familiar rate constant typically used in a master equation description of the same system. Taken together, these considerations make it possible to convert between a master equation description in terms of molar concentrations, and the kinetic Monte Carlo framework in terms of discrete molecular population sizes.

The kinetic Monte Carlo algorithm

In a kinetic Monte Carlo simulation, the system is evolved by a single chemical reaction event at every time step. Each time step involves the selection of a reaction and a time interval since the last reaction occurred. The chosen reaction is then implemented by adjusting the populations of reactant and product molecules, and the total system time is updated according to the chosen reaction time interval. The time interval and reaction are randomly selected according to the total propensity of all reactions in the system, using the following method.

[1] The propensity is the probability per unit time that a particular event occurs.

[2] For molar concentrations, the system size must also take into account Avogadro's number, N_A.

[3] For example, there are $n = 2$ molecules involved in the bimolecular reaction $A + B \xrightarrow{\mu} C$.

Supposing that the next reaction occurs during the time interval $(\tau + t, \tau + t + dt]$, where τ is the current total system time. The waiting time t before this reaction happens is then determined by the probability that no reactions at all occur in the interval $[\tau, \tau + t]$, giving

$$t = -\frac{1}{k} \ln (r_{\mathrm{t}}) \tag{2.1}$$

where K is the sum of all propensities, r_{t} is a random number in the interval $(0, 1]$, and it has been assumed that all events have Poisson distributed waiting-times. Compared to a simulation in which the time interval between consecutive updates is fixed, this algorithm has the advantage that it automatically jumps ahead in time according to the reaction propensities. Thus, no computational time is wasted simulating periods when no reactions can occur.

The reaction J, which occurs after this waiting time, can now be determined by considering the propensity of each reaction k_i relative to that total propensity of all reactions K. The sum of propensities up to a given reaction j can be written as $R_j = \sum_{i=0}^{j} k_i$, such that $R_\infty = K$. Using this, and a random number r_{r} in the interval $(0, K]$ such that $r_{\mathrm{r}} = R_J$, reaction J can be identified from the condition $R_{j-1} < r_{\mathrm{r}} \leq R_{j+1}$. This reaction is then used to adjust the populations of each species involved.

Once the reaction has been performed and the system time update, the propensities for all reactions are recalculated and the simulation moves onto the next time step. In this way, every reaction with a non-zero propensity has a finite probability of occurring, but those with higher rate or larger number of particles will have a higher probability and on average will be chosen more often.

Implementation

Using this algorithm, I implemented various kinetic models of amyloid fibril self-assembly in `FORTRAN95`, with random numbers generated using the well known and robust Mersenne-Twister algorithm. In these simulations, the chemical species involved were either monomers, or fibrils of a given length.

All simulation parameters were input in SI units from which molar concentrations were converted into numbers of particles according to the chosen system volume. No constraint was placed on the maximum length of an individual filament or the number of filaments produced other than to conserve the total number of monomers in the system. Typical runs contain 150 replicates for each set of parameters values.

Data was output for growth curves and distributions of filament lengths in terms of the number of monomers in a particular state; i.e. free, or within filaments of a given range of lengths. Growth curves were stored whenever the total monomer concentration changed by $\pm 0.5\%$, and the filament length distributions were written upon reaching specific fractions of the total monomer concentration, such as $0.5 m_{\mathrm{tot}}$.

2.2 Averaging the kinetic Monte Carlo simulation data

In order to compare kinetic Monte Carlo simulation results to the mean-field prediction, I compute the mean behaviour of the simulation results as follows.

Growth curves

By virtue of the kinetic Monte Carlo approach, replicate simulation runs do not have neatly coinciding time intervals; however, the fraction of monomers incorporated into filaments can (and do) have intervals which coincide between replicates. Moreover, as will be seen in later chapters, the lag phase and subsequent growth are often governed by vastly different time scales, meaning that replicates appear identical but significantly translated (rather than scaled) in time.

Hence, I constructed the mean growth curve from a set of replicates by considering the mean time at which they all reached a particular fraction of the total protein concentration, for each fraction of the total protein concentration recorded [(4)]. Thus, for each fraction of the total protein concentration, the mean growth curve describes the mean time at which that fraction was observed. This trivially provides the mean lag-time τ_{ϕ_M}, which is the mean of the times that each replicate reached the threshold fraction of the total monomer concentration ϕ_M.

To find the mean and variance of the maximum aggregation rate, I used the method presented below in §2.3.2 to extract the maximum aggregation rate from each replicate growth curve.

Length distributions

Filament length distributions were obtained from kinetic Monte Carlo simulations at specific events of interest; for example, reaching the fraction of aggregated material that defines τ_{ϕ_M}. For each of these I determined the average length distribution by constructing a histogram covering the entire range of filament lengths contained in the replicate data sets, with equal bin spacing. The mean and standard deviation of the number of filaments with lengths that lie within each bin was then calculated from the replicate data. In this way, I capture the mean number of filaments for a given range of lengths, and also highlight the extent of fluctuations between replicate simulations with the same fraction of aggregated material.

2.3 Extracting model parameters from experimental data

A key part of the work in this thesis involves fitting the analytic expressions of a given model for fibril aggregation kinetics, to experimental and simulation results thought to be described by that model. To do this, I make use of the Levenberg-Marquardt algorithm in the context of least squares fitting.

(4) In contrast to the mean fraction of the total protein concentration at a specific time, which is the way that the average of replicate experimental growth curves is typically found.

2.3.1 The Levenberg-Marquardt algorithm

The Levenberg-Marquardt algorithm is commonly used to find a "best fit" between an analytic expression and data which the expression is thought to describe. More generally it provides a fast method for finding the minimum value of a scalar function by adjusting the parameters on which the function depends; hence, in this case the function being minimised is Chi-squared (χ^2), which is the sum of the squared residuals for a given analytic expression and dataset, weighted by the confidence in each data point.

The algorithm finds a minimum value of χ^2 by alternating between the Gauss-Newton algorithm and the method of steepest descent. The Gauss-Newton algorithm is a reliable method for non-linear least-squares minimisation. Faster and more robust than Newton's method for finding the minimum value of a scalar function, the Gauss-Newton algorithm quickly converges to a local minimum but takes many small steps to reach it. To speed up the process, the method of steepest descent is employed when there is a steep local gradient in χ^2 with respect to any of the parameters [5]. Thus the Levenberg-Marquardt algorithm takes large steps down steep gradients in parameter space and slower, more accurate steps as a minimum is approached.

The Levenberg-Marquardt algorithm is sometimes criticised for finding local minima, and different minima when started with different initial parameter values. While these can be problematic, the algorithm's speed means that a large portion of parameter space can be quickly mapped. As a result, even if most of the minima found turn out to be local minima, a picture can be built up of where the global minimum actually is.

2.3.2 Determining the lag-time and maximum aggregation rate from growth curves

From experimental and simulation growth curves, I determined the maximum aggregation rate by taking the first derivative $\frac{d(M(t)/m_{\mathrm{tot}})}{dt}$ of the normalised data

[5] Where a small change to one or more of the parameters in the analytic expression yields a large decrease in χ^2.

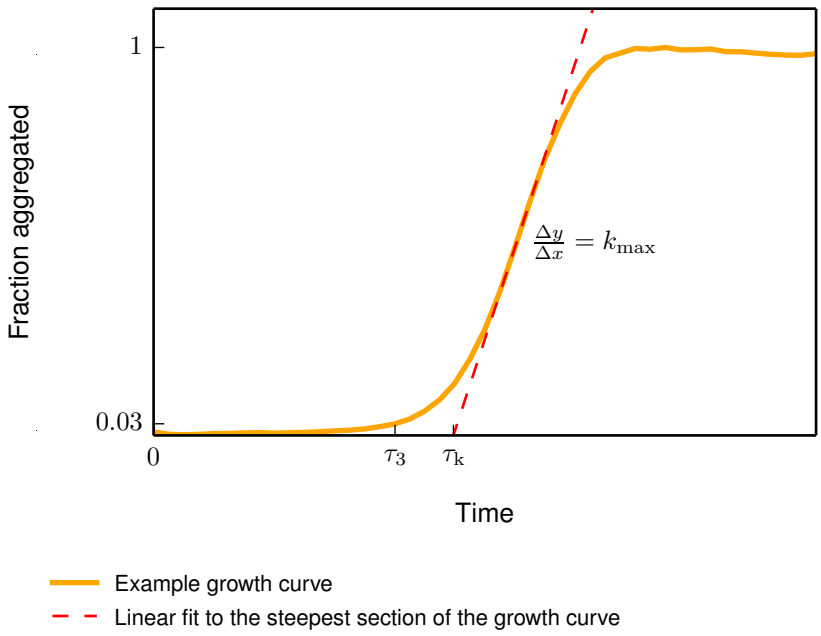


Figure 2.1 Illustration showing how the lag times τ_k and τ_3 (i.e. when $\phi_M = 0.03$), and the maximum aggregation rate k_{max} are determined from a growth curve; the dashed line is the best fit of $M(t)/m_{tot} = k_{max}(t - \tau_k)$ to the steepest part of the growth curve.

and finding the maximum $\tilde{k}_{max}$. This value should be the same as the actual maximum aggregation rate k_{max} illustrated in figure 2.1, but local fluctuations in the data mean that this is not guaranteed. For this reason, I obtain a better estimate using the following method.

First, I extracted the values of $M(t)/m_{tot}$ and t for which $\frac{d(M(t)/m_{tot})}{dt} \geq 0.75\tilde{k}_{max}$: this data encompass a large portion of the approximately linear section of the growth curve around the inflection point. By fitting this data to the line described by

$$\frac{M(t)}{m_{tot}} = k_{max}(t - \tau_k)$$

illustrated by the dashed line in figure 2.1, the values of k_{max} and τ_k can be found.

The lag time defined by the time to reach a threshold fraction of aggregate, was found simply by seeking the value of t for which $M(t)$ first crosses M_ϕ. This I did by iterating over the growth curve data and linearly interpolating between the two values of $M(t)$ closest to ϕ_M.

Extracting the lag-time scaling exponent γ

The lag time is expected to scale with total protein concentration according to the expression

$$\tau_{\mathrm{lag}} = A m_{\mathrm{tot}}^{-\gamma} \tag{2.2}$$

where A is an arbitrary scale factor. Thus, the lag-time scaling exponent γ was determined by fitting the lag times in a given range of total protein concentrations (typically $10 \leq m_{\mathrm{tot}} < 750$ μM) to the linear expression

$$\log\left(\tau_{\mathrm{lag}}\right) = \log\left(A\right) - \gamma \log\left(m_{\mathrm{tot}}\right)$$

Chapter 3

The effects of stochasticity on two autocatalytic polymerisation models

3.1 Introduction

In order to understand the variability in the experimental growth curves in figure 3.1 it is necessary to move away from the mean-field description of Cohen et al. [12] (see §1.3.2). To do this, I make use of kinetic Monte Carlo (kMC) simulations, which provide a way to resolve discrete reaction events and thus probe the stochastic fluctuations that arise from randomness in the timing and order of individual molecular reactions.

In this chapter I will compare the effects of stochasticity on the autocatalytic polymerisation models of fragmentation and heterogeneous nucleation (described in §1.3.2). This will be done using the reactions for primary nucleation, elongation, and autocatalysis from table 1, implemented in the kinetic Monte Carlo simulation framework described in §2.1. The results of these simulations will also be compared to the mean-field predictions for these models and to experimental data, both of which reveal interesting behaviour which will be investigated in subsequent chapters.

First, however, I must establish realistic values for the rate constants involved in these models.

3.2 Estimating the rate constants from experimental data

Estimated values for the rate constants of primary nucleation (k_{n}), elongation (k_{+}), and fragmentation (k_{f}), were determined from the parameter values of the mean-field predictions in §1.3.4, by fitting to experimental data. As will be seen below, the values found appear physically realistic but are unlikely to be the true rates of the physical processes, due to the limitations of the data available. These estimated values simply form a baseline, providing information about the relative contribution of each mechanism to the aggregation process in one case so that deviations from this behaviour can be understood in terms of changes to the model parameters.

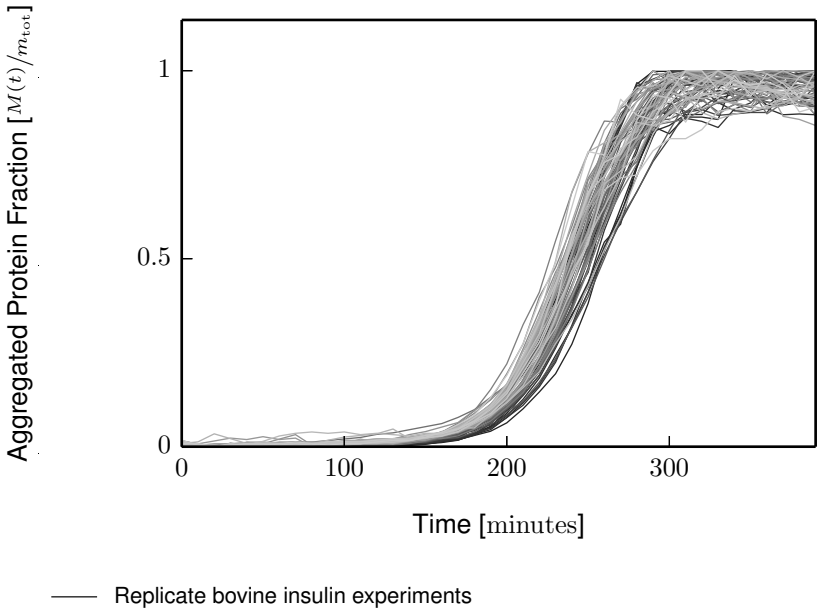


Figure 3.1 Replicate growth curves for the self-assembly of amyloid fibrils from 131 μM bovine insulin, previously shown in figure 1.3. These 68 replicates were obtained simultaneously, by my collaborator Ryan Morris [74], and have been normalised by the maximum fluorescence of each curve.

Replicate experimental growth curves for different concentrations of bovine insulin in the range $17 - 698$ μM were averaged by finding the mean of the normalised fluorescence signal at each time point. To the average growth curve of each protein concentration, the analytic expression for the mean-field growth curve (equation 1.14) was then fitted using the least squares minimisation technique presented in §2.3.1; this gave values of the parameters κ and δ (defined in §1.3.3). An example fit produced by this process is shown in figure 3.2 for the 131 μM bovine insulin data shown in figure 3.1.

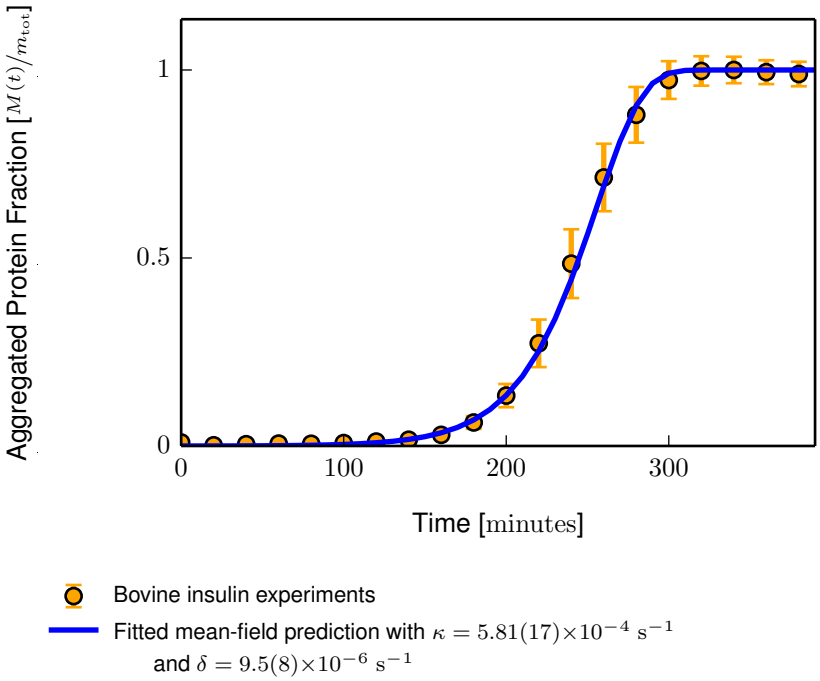


Figure 3.2 An example fit of equation 1.14 to experimental data; in this case, the mean normalised fluorescence value at each time point of the replicate growth curves shown in figure 3.1 for 131 μM bovine insulin, with error bars indicating the standard deviation. These experiments were not "seeded" (i.e. $M(0) = 0$ and $N(0) = 0$) and so κ and δ are the only fitted parameters.

The fitted parameters κ and δ each consist of a combination of rate constants and the total protein concentration. In the case of κ, this combination depends on which autocatalysis model is being considered. Inspection of the scaling behaviour of the average lag time and maximum aggregation rate with total protein concentration [(1)] indicates that this experimental data is best described by the fragmentation model only, with no evidence of heterogeneous nucleation

[(1)] This will be shown below in figures 3.7 and 3.8, respectively.

being involved. Thus, I can obtain an estimate for k_f from the experimental data, and I am then free to define k_h such that the two models have the same κ at a particular protein concentration and, hence, a similar maximum aggregation rate and lag time.

In order to extract the rate constants from κ and δ, additional constraints are needed. These constraints are provided by additional experimental data, such as the mean lag time and mean fibril length; however, the nucleus size n_c has not accurately been determined experimentally for any amyloid fibril formation process, and so I have assumed $n_\mathrm{c} = 2$ for convenience.

These experiments were not "seeded" with preformed fibrils and so I can additionally assume that $M(0) = 0$ and $N(0) = 0$, such that the lag times provide an estimate of the ratio δ/κ; i.e. the ratio $k_\mathrm{n}/k_\mathrm{f}$.

The mean fibril length, obtained from electron micrographs by Morris et al. [74] after aggregation had completed, provides $\overline{L}(\infty) = \kappa/k_\mathrm{f}$, where each bovine insulin "monomer" contributes approximately 4 Å to the fibril length. This fixes k_f and allows k_+ to be determined from κ. There is no way to separate this elongation rate into monomer addition and depolymerisation rates, and so I have assumed $k_- = 0$ for the fragmentation model; as will be seen shortly, a non-zero depolymerisation rate is necessary for the heterogeneous nucleation model.

Encouragingly, very similar values and confidences for $\kappa/\sqrt{m_\mathrm{tot}}$ [(2)] were found by fitting the average growth curves of total protein concentrations below approximately 175 μM, and these values were robust against the choice of starting values for the fitted parameters. For reasons which will become clearer in §6, higher protein concentrations give a lower value of κ than expected, but these fits were still robust against the choice of starting values.

The value of δ, on the other hand, was less consistent: its value turns out to be far less important to the shape of the growth curve. This is because autocatalysis and primary nucleation are both contributing to the lag phase and cannot simply be decoupled by inspecting the growth curve; at best, δ can be said to have a value significantly smaller than κ. For this reason, the effects of different values of the primary nucleation rate will be dealt with in §5.

(2) This is the concentration independent form of κ for the fragmentation model.

Taking account of these considerations and the fact that the rate constants appear as products within the fitted parameters, it is not possible to determine the exact reaction rates using this method; however, the values obtained, do appear physically realistic. In the case of bovine insulin, I find that $k_+ = 5{\times}10^4\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$, $k_\mathrm{f} = 3{\times}10^{-8}\ \mathrm{s}^{-1}$, and $k_\mathrm{n} = 3{\times}10^{-7}\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$ produce results consistent with the experimental data [3].

To allow comparison between the fragmentation and heterogeneous nucleation models of autocatalysis, I now determine an appropriate value for the heterogeneous nucleation rate using the two definitions of κ (equations 1.17 and 1.18): $k_\mathrm{h} = {k_\mathrm{f}\,(n_\mathrm{s}!)}/{m^{n_\mathrm{s}}} = 24\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$, when $m = 50\ \mu\mathrm{M}$ and $n_\mathrm{s} = 2$. To account for the loss of monomers that happens through fragmentation at filament ends, the heterogeneous nucleation model additionally requires a non-zero depolymerisation rate: in the case where $n_\mathrm{c} = 2$ this is $k_- = k_\mathrm{f} = 3{\times}10^{-8}\ \mathrm{s}^{-1}$, but for higher n_c, depolymerisation cannot properly account for fragmentation of filaments of length $1 < i < n_\mathrm{c}$ in these simulations due to the differences between reactions 1.7 and 1.9.

Since the kinetic Monte Carlo framework considers individual particles in order to resolve discrete reaction events, the rate constants determined above, and all concentrations of protein, must be converted into discrete numbers of particles according to a system volume V, as described in §2.1. Hence, at smaller volumes the effects of stochasticity are expected to increase due to the reduced number of particles. This will be seen to be particularly important during the lag phase when the number of filaments is expected to be very small. For reasons which will become clear shortly (in §3.3.3), I use a volume of $V = 0.83$ pl throughout this thesis unless otherwise stated.

[3] The primary nucleation rate should be expressed in units of $\mathrm{M}^{-(n_\mathrm{c}-1)}\ \mathrm{s}^{-1}$, unless n_c is defined; for most of this thesis I will assume $n_\mathrm{c} = 2$.

3.3 Comparison of kinetic Monte Carlo simulations with the mean-field predictions

3.3.1 The stages of autocatalytic polymerisation

In order to add context to the discussion in other chapters, I first present an overview of the mechanisms which dominate at specific points in the self-assembly kinetics.

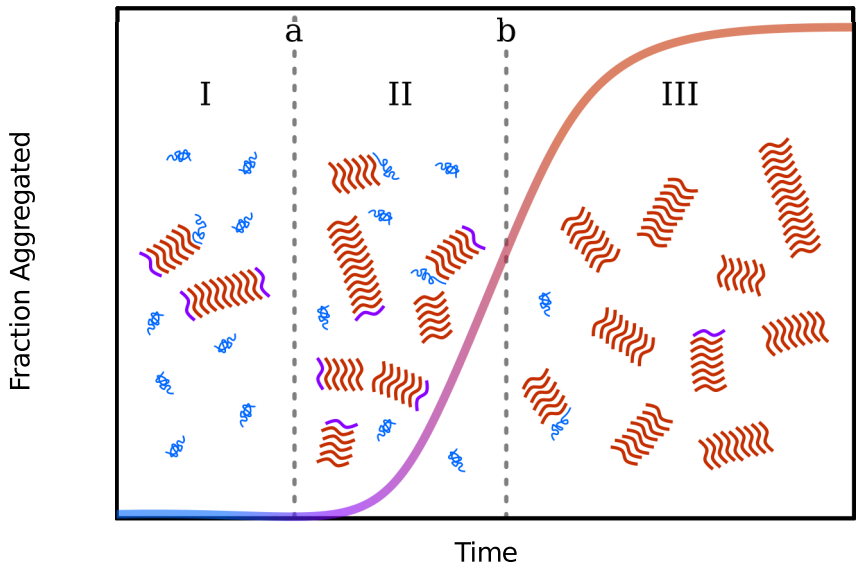


Figure 3.3 A cartoon depicting the mechanisms that dominate different stages of the sigmoidal growth curve in the fragmentation model for amyloid fibril self-assembly. At early times (**I**), few filaments are present and elongation dominates. Once this has incorporated a sufficient quantity of protein into fibrils (**a**), autocatalysis takes over during the exponential growth phase (**II**), up to the point where the aggregation rate reaches its maximum (**b**). After this, the creation of more filaments through autocatalysis cannot balance the depletion of monomers by elongation, causing growth to slow and become elongation dominated again (**III**); continued fragmentation during this stage, leads to a large number of short filaments being present in the long-time limit. The behaviour of the heterogeneous nucleation model would differ at point **b**, which would be shifted to earlier times, and in stage **III**, where fewer and longer filaments would be present in the long-time limit.

The growth curves for either model of autocatalysis can be considered in three stages, delimited by: (a) the end of the quiescent phase and (b) the point at which the aggregation rate is maximal. These are illustrated in figure 3.3.

During the quiescent phase (**I** in figure 3.3), the low rate of primary nucleation results in few filaments being produced. These grow rapidly due to the abundance of free monomer, but since they are so few in number the total rate of aggregation is low. Once a filament reaches a significant length, autocatalytic processes become important, leading to the formation of more growth-competent filaments.

Once autocatalysis starts to become important, both the number of filaments and their combined protein content will grow exponentially while sufficient monomer is available (**II** in figure 3.3). Thus the end of the lag phase is marked by a strong increase in the number of filaments. The exponential growth phase continues until depletion of the monomer population starts to arrest growth more rapidly than the formation of new filaments accelerates it; this crossover happens at the maximum aggregation point. Following this point, aggregation slows and ultimately stops once an equilibrium is reached between polymerisation, and either depolymerisation or the fragmentation of filaments into pieces shorter than n_c (**III** in figure 3.3).

3.3.2 The effects of stochasticity on the fragmentation and heterogeneous nucleation models

Using the rate constants obtained in §3.2 as a starting point, I performed kinetic Monte Carlo simulations of the fragmentation and heterogeneous nucleation models. For the fragmentation model, reactions for primary nucleation (reaction 1.6), the disintegration of filaments smaller than the primary nucleus (reaction 1.7), polymerisation (reaction 1.8), and breakage of a filament (reaction 1.10) were used. In the heterogeneous nucleation model, reactions 1.6, 1.7 and 1.8 were also used, along with those for depolymerisation (reaction 1.9) and heterogeneous nucleation (reaction 1.11).

The replicate growth curves that I obtained from these simulations are shown in figures 3.4 and 3.5 for the fragmentation and heterogeneous nucleation model, respectively. Here, I also show the average growth curve for the set of replicates, and the mean-field prediction given in equation 1.14. From these it is obvious that the average of the replicate simulations of both models do not agree with the mean-field prediction.

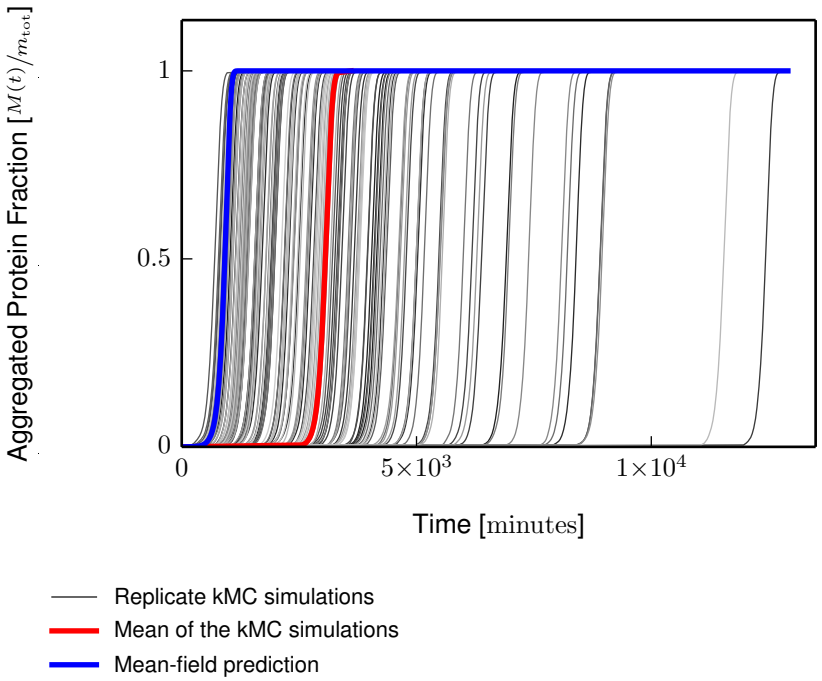


Figure 3.4 Growth curves produced by kMC simulations of the fragmentation model for amyloid fibril self-assembly, along with the mean-field prediction given in equation 1.14, where the parameter values were $m_{\mathrm{tot}} = 10\ \mu\mathrm{M}$, $n_{\mathrm{c}} = 2$, $N(0) = 0$, $k_{\mathrm{n}} = 3\times10^{-7}\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$, $k_{+} = 5\times10^{4}\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$, $k_{-} = 0$, $k_{\mathrm{f}} = 3\times10^{-8}\ \mathrm{s}^{-1}$, and $V = 83\times10^{-14}\ \mathrm{l}$. The average of the simulations was obtained from the 150 replicates shown, using the method in §2.2.

In figure 3.4, the average growth curve of the simulations and mean-field prediction both have the same shape, but the former has a longer lag phase. This is due to the difference of time scale for events during the lag phase, and those later in the growth curve, resulting in apparently identical growth curves with vastly different lag times. The mean-field description given by equation 1.14 correctly accounts for the faster time scale of autocatalytic polymerisation, but not the slower time scale of discrete primary nucleation; hence, it can reproduce

the shape of the growth curve, but not the mean lag-time. Thus, the average growth curve of the simulation and mean-field prediction would agree if the correct mean-field lag time were added onto the latter, as will be shown in §5.

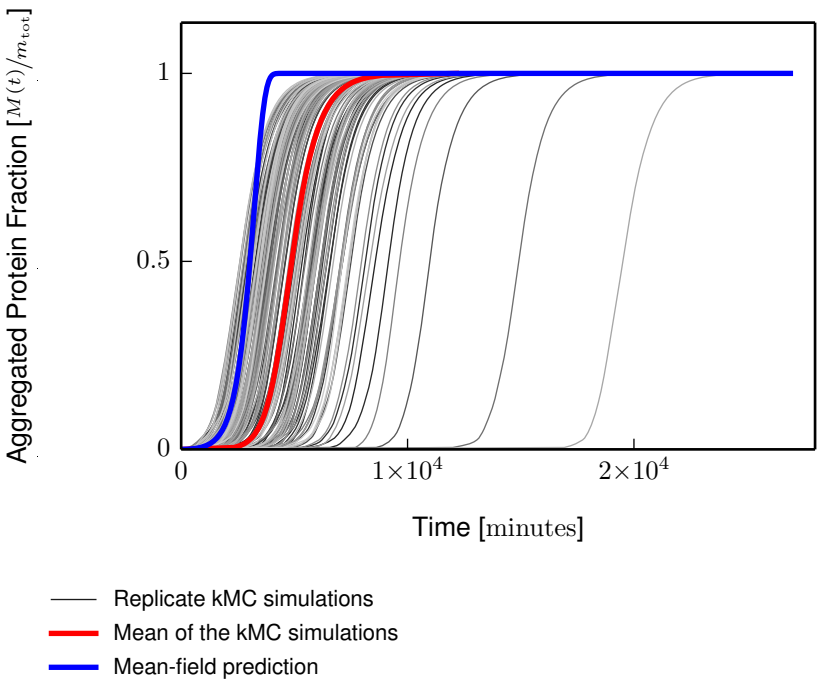


Figure 3.5 Growth curves produced by kMC simulations of the heterogeneous nucleation model for amyloid fibril self-assembly, along with the mean-field prediction given in equation 1.14, where the parameter values were $m_{\text{tot}} = 10$ μM, $N(0) = 0$, $n_{\text{c}} = 2$, $k_{\text{n}} = 3{\times}10^{-7}$ M^{-1} s^{-1}, $k_{+} = 5{\times}10^{4}$ M^{-1} s^{-1}, $k_{-} = 3{\times}10^{-8}$ s^{-1}, $k_{\text{h}} = 24$ M^{-1} s^{-1}, $n_{\text{s}} = 2$, and $V = 83{\times}10^{-14}$ l. The average of the simulations was obtained from the 150 replicates shown, using the method in §2.2.

The same observations hold for the growth curves produced by the heterogeneous nucleation model in figure 3.5. However, here the average maximum aggregation rate of the simulations — and indeed the shape of the growth curve above $0.5m_{\text{tot}}$ — also appears quite different to the mean-field prediction. This discrepancy arises because the analytic solution in this form does not take full account of monomer depletion and its impact on the heterogeneous nucleation mechanism. Higher order corrections to the this solution, provided by Cohen et al. [12], do capture these effects, but the increased complexity makes them impractical for use in extracting parameters from experiential data. However, as the experimental data appears best explained by the fragmentation model, I will have need of these more complex solutions.

3.3.3 The effects of system volume, primary nucleation and seeding

I stated that the poor agreement between the average growth curve of the kinetic Monte Carlo simulations and the mean-field prediction was due to the primary nucleation process. This I will now illustrate by "seeding" the kinetics with a number of preformed filaments prior to self-assembly, and in doing so, explain the choice of simulated volume $V = 0.83$ pl.

The concentration of filaments present early on in the kinetics, determine the rate of aggregation; hence, a high concentration will reduce or bypass the lag phase, while a low concentration will allow autocatalysis (and primary nucleation) to contribute to the formation of more filaments. The growth curves which were shown in figures 3.4 and 3.5 where the rate of primary nucleation was assumed to be low ($\delta \ll \kappa$) are expected to be in the latter of these two conditions, and indeed display an notable lag phase.

In these kinetic Monte Carlo simulations, the primary nucleation and autocatalysis processes always produce one filament per event, regardless of system volume; thus, the choice of system volume dictates the relative contribution which this one new filament makes to the concentration of all filaments. Hence, a single filament in a small volume constitutes a "high concentration", and from the above argument, if the simulation volume were too small, then a single nucleation event would bypass the exponential growth phase completely, even if the primary nucleation rate was very low.

As explained in §2.1, the system volume also affects the propensities of each event. In the kinetic Monte Carlo simulation framework, it is assumed that every event can be described by a Poisson process with a rate that scales inversely with the concentration of particles produced. This is simply because the process of producing a high concentration of a given species of particle requires more reaction events than would the production of a lower concentration of the same species in the same volume. Consequently, the rate of production of a high concentration is slower than a lower concentration, even if both are treated as single events.

Using primary nucleation as an example, the Poisson rate at which nuclei form can be written as $\eta/N(0)$, where $\eta = k_n m_{tot}^{n_c}/n_c!$ is the propensity given in reaction 1.6 and $N(0) = M(0)/n_c = 1/N_A V$ is the number concentration of nuclei produced. From this, the rate at which nuclei form can be seen to scale with system size; in smaller systems the waiting-time distribution is broader and nuclei are formed more slowly. This particular effect will be revisited in §4 and §5.

A large portion of the variability in the lag phase of the growth curves in figures 3.4 and 3.5 can thus be explained by the width of the waiting-time distribution for primary nucleation events at this particular volume, $V = 0.83$ pl. This I will now illustrate by comparison of those growth curves with the ones in figure 3.6b. These were produced using the same parameters as those in figure 3.4 for the fragmentation model, but with primary nucleation disabled and the simulations seeded with $N(0) = 1/N_A V$ filaments, as indicated. Here, the agreement between the average of the simulation growth curves and the mean-field prediction is improved, compared to the case where primary nucleation was active. However, it is now apparent that the mean-field prediction also fails to account for monomer depletion at later times in the fragmentation model.

The growth curves in figure 3.6 were obtained for different values of $N(0)$ by altering V in the range $83\times10^{-16} - 83\times10^{-13}$ l, but always starting with a single filament. From the way that the lag-time decreases with decreasing volume in these figures, it can be seen that when $k_n = 0$ and $N(0) > 0$, in order for the kinetics to produce the same mean lag-time as the experimental data at the same protein concentration, the minimum system volume must be $V = 1/N_A N(0)$, according to the lag-time equation 1.21, where $N(0) = 2\times10^{-12}$ M; hence giving $V = 0.83$ pl.

The "kick" in the lower part of the growth curves in figure 3.6d is due to the concentration of nuclei being slightly below the threshold at which additional filament ends would make no appreciable difference to the rate of aggregation. Consequently, a lower aggregation rate persists until the first autocatalysis event generates extra growth-competent ends and boosts the rate of aggregation. This occurs at a random time in each replicate and illustrates how a single rare event can have a profound affect on the growth curve.

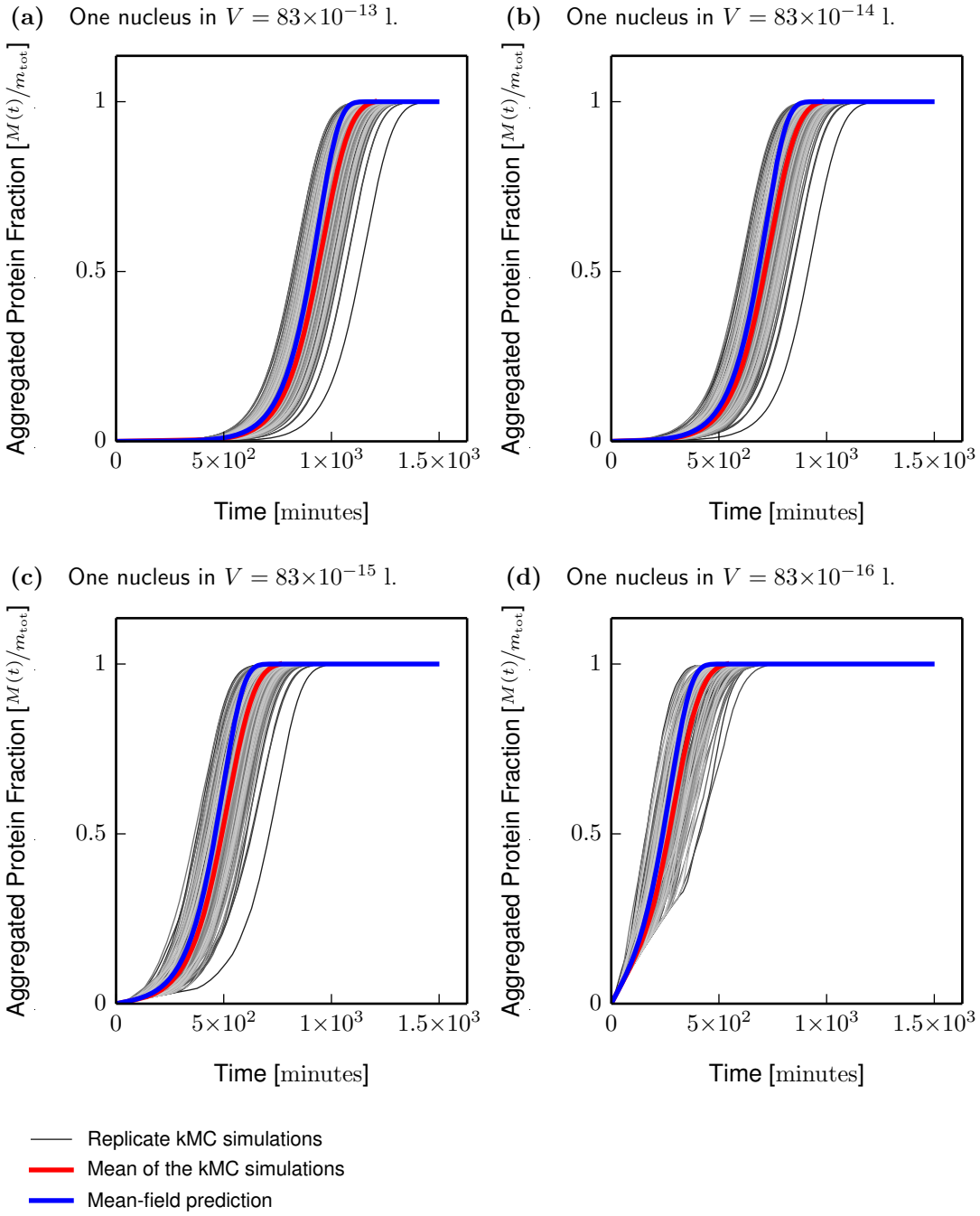


Figure 3.6 Comparison of growth curves produced at different system volumes (V) when self-assembly is "seeded" with a single filament ($N(0) = {}^{1}/_{N_{\mathrm{A}}V}$) at $t = 0$ and primary nucleation is disabled ($k_{\mathrm{n}} = 0$); where the common parameter values were $m_{\mathrm{tot}} = 10$ μM, $n_{\mathrm{c}} = 2$, $k_{+} = 5\times10^{4}$ M^{-1} s^{-1}, $k_{-} = 0$, and $k_{\mathrm{f}} = 3\times10^{-8}$ s^{-1}. Each plot shows 150 replicate growth curves produced by kMC simulations of the fragmentation model for amyloid fibril self-assembly, along with the mean-field prediction given in equation 1.14, and the average of the replicates (obtained using the method in §2.2).

The equivalent growth curves for heterogeneous nucleation are not shown because the effects are almost identical and do not add to the understanding of the processes involved.

3.3.4 The scaling of maximum aggregation rate with total protein concentration

For the autocatalysis model, the maximum aggregation rate $k_{\max}$ is predicted to scale with the total protein concentration according to κ. This takes the form of a power law: for fragmentation $\kappa \sim m_{\text{tot}}^{1/2}$, while heterogeneous nucleation with secondary nucleus size n_{s}, predicts $\kappa \sim m_{\text{tot}}^{(n_{\text{s}}+1)/2}$.

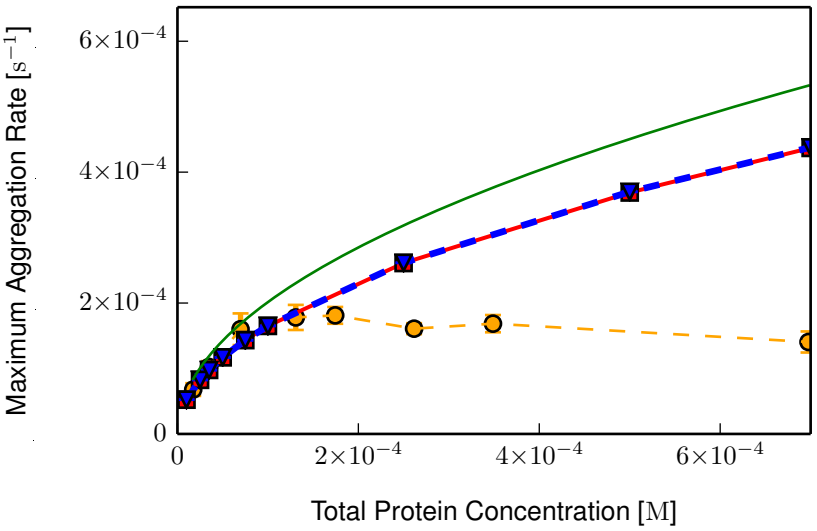
Figure 3.7 shows these predicted scaling behaviour, along with the kinetic Monte Carlo simulation results for both the fragmentation and heterogeneous nucleation models. While the trend is broadly the same, the simulations depart from the prediction at higher total protein concentrations and more strongly in the case of the heterogeneous nucleation model. This is likely to be caused by the simulations containing fewer filaments than is predicted in the mean-field limit, which means that the increase in monomer abundance does not result in the expected increase in aggregation rate.

Also shown in figure 3.7a are the experimentally determined maximum aggregation rates for bovine insulin. The large difference between these and the prediction at high protein concentrations appears to be caused by a similar limitation of the number of filaments, and will form the subject of §6.

3.3.5 The scaling of lag time with total protein concentration

For the autocatalysis model, the lag-time equation 1.22 makes the strong prediction that the lag time will scale with total protein concentration according to $\tau_{\text{lag}} \sim m_{\text{tot}}^{-\gamma}$ for $\gamma = {}^{(n_{\text{s}}+1)}/_{2}$, where n_{s} is the secondary nucleus size. For the fragmentation model where $n_{\text{s}} = 0$, this gives $\gamma = {}^{1}/_{2}$, whereas for the heterogeneous nucleation model, $\gamma = {}^{3}/_{2}$ when $n_{\text{s}} = 2$.

(a) The fragmentation model where $k_- = 0$ and $k_\mathrm{f} = 3{\times}10^{-8}\ \mathrm{s}^{-1}$.



(b) The heterogeneous nucleation model where $k_- = 3{\times}10^{-8}\ \mathrm{s}^{-1}$, $k_\mathrm{h} = 24\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$, and $n_\mathrm{s} = 2$.

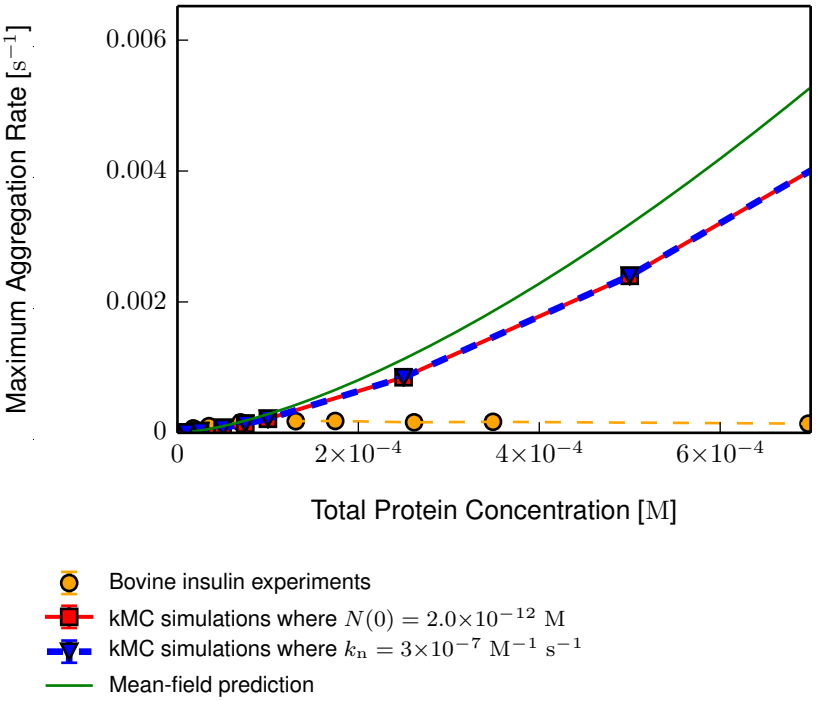
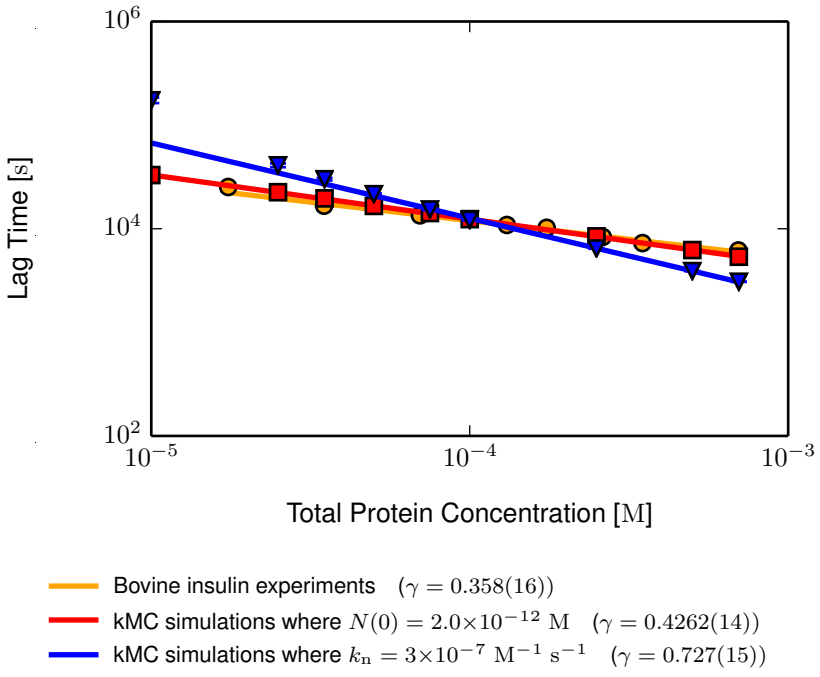


Figure 3.7 The scaling behaviour of k_max with m_tot in **(a)** the fragmentation model, and **(b)** the heterogeneous nucleation model, where the common parameter values were $n_\mathrm{c} = 2$, $k_+ = 5{\times}10^4\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$, and $V = 83{\times}10^{-14}$ l. Each kMC simulation point is the mean of 150 replicates with error bars indicating the standard deviation (method in §2.2); the lines are to help differentiate the data sets. For comparison, the experimental data for bovine insulin and mean-field prediction (equation 1.19) for the value of κ specific to each mode, are also shown.

(a) The fragmentation model where $k_- = 0$ and $k_f = 3\times10^{-8}\ s^{-1}$.



(b) The heterogeneous nucleation model where $k_- = 3\times10^{-8}\ s^{-1}$, $k_h = 24\ M^{-1}\ s^{-1}$, and $n_s = 2$.

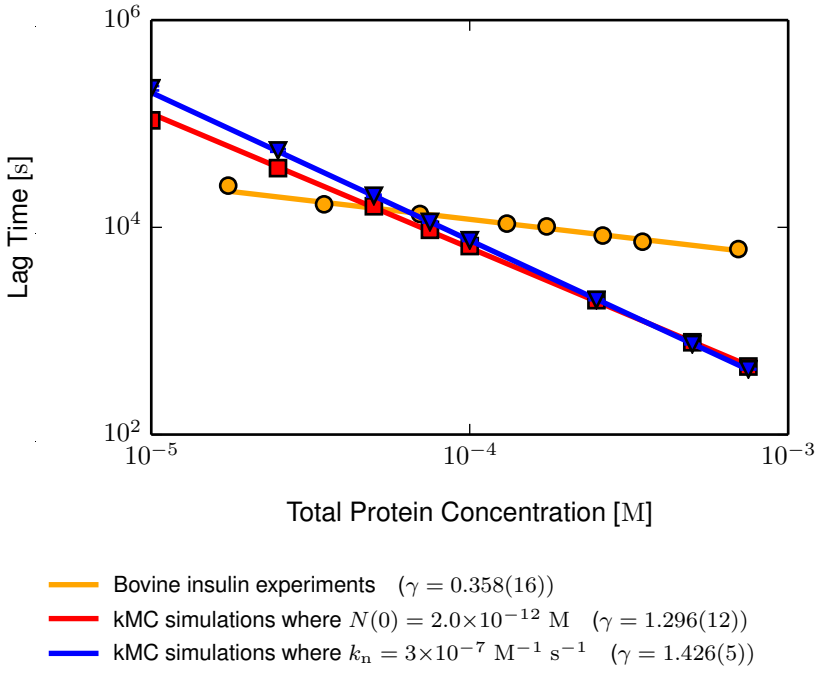


Figure 3.8 The scaling behaviour of τ_3 with m_{tot} in **(a)** the fragmentation model, and **(b)** the heterogeneous nucleation model, where the common parameter values were $n_c = 2$, $k_+ = 5\times10^4\ M^{-1}\ s^{-1}$, and $V = 83\times10^{-14}$ l. Each kMC simulation point is the mean of 150 replicates with error bars indicating the standard deviation (method in §2.2). The experimental results for bovine insulin are shown for comparison. Each line indicates the best fit of the power law $\tau_{lag} \propto m_{tot}^{-\gamma}$ to the data set (method in §2.3.2).

The lag times that I obtained from both autocatalytic models with and without seeding, are shown in figure 3.8, along with the experimental data previously shown in figure 1.8. As I mentioned in §3.2, the experimental data appears well reproduced by the seeded fragmentation model, and the value of γ obtained for these, and for the heterogeneous nucleation model also in the absence of primary nucleation, agrees very well with the scaling behaviour predicted by equation 1.22; i.e. $\gamma = 1/2$ for the fragmentation model, and $\gamma = 3/2$ for the heterogeneous nucleation model for $n_s = 2$.

However, primary nucleation clearly increases the scaling exponent of both autocatalytic models. Moreover, the scaling of lag time with protein concentration appears no longer to resemble a power law at low protein concentrations. The origin of this departure, or "kink", in the lag-time scaling behaviour, and the increased scaling exponent in the presence of primary nucleation, will be the subject of §5. For now, however, it should simply be noted that stochastic effects of the primary nucleation process causes a departure from the mean-field predictions at early times.

3.3.6 Evolution of the filament length distribution

As I have shown, stochastic effects can have a considerable impact on the growth curves, and so it is important to determine whether they also influence the distribution of filament lengths.

The mean-field description of autocatalytic polymerisation (see §1.3.2) provides predictions for the distribution of filament lengths at both the long-time steady state, and at earlier times when the filament population is in equilibrium with the free monomer population. Cohen et al. [13] predicted that the shape of these length distributions, at both points in the kinetics, would be different in both models due to the difference in the way that fragmentation and heterogeneous nucleation generate new filaments (see §1.3.4).

In the fragmentation model, each monomer junction along the length of a filament has an equal probability of breakage; thus, the total probability that a given filament breaks increases as the filament becomes longer. Consequently, under the action of fragmentation, the length distribution is expected to collapse towards shorter filaments, and fragments containing fewer than n_c monomers will disintegrate into free monomer. In the long-time steady state, this results in a narrow skew normal distribution with a peak close to n_c. At earlier times, while sufficient free monomer is available, this is prevented by the action of elongation increasing the length of all filaments, resulting in a broader skew normal distribution.

In the heterogeneous nucleation model, new filaments are produced with a fixed length n_s. These become longer through the action of elongation, and hence the number produced also increases due to autocatalysis. As the new filaments grow, so more are created by heterogeneous nucleation. This positive feedback results in a Poisson distribution. However, once the monomer population is consumed by elongation and heterogeneous nucleation, depolymerisation will gradually cause the distribution to collapse into an exponential distribution in the long-time steady state.

Using the kinetic Monte Carlo simulation data, I will now show that another, distinct distribution exists at times between the dynamic equilibrium at early times and the long-time steady state, for both models; this I shall refer to as the "end-point" distribution. In the fragmentation model, these distributions scale according to the ratio $2m_{\mathrm{tot}}k_+/k_{\mathrm{f}}$ [(4)], and are otherwise independent of the rate constants and total protein concentration. For this reason I have normalised the filament lengths by the predicted mean (equation 1.25 with the appropriate κ for the model), making it easier to compare between the two models.

Filament length distributions produced by the fragmentation model are shown in figure 3.9 for the three indicated points in the kinetics. At early times and around the point where the aggregation rate is maximal, the distributions produced by the kinetic Monte Carlo simulation match the mean-field prediction described above. The end-point exponential distribution depicted in figure 3.9d is not the steady-state distribution, but is instead an intermediate state between the peaked distribution at earlier times and the much narrower, peaked distribution at later

(4) The equivalent for the heterogeneous nucleation model is $2m_{\mathrm{tot}}^{1-n_s}k_+/k_{\mathrm{h}}$.

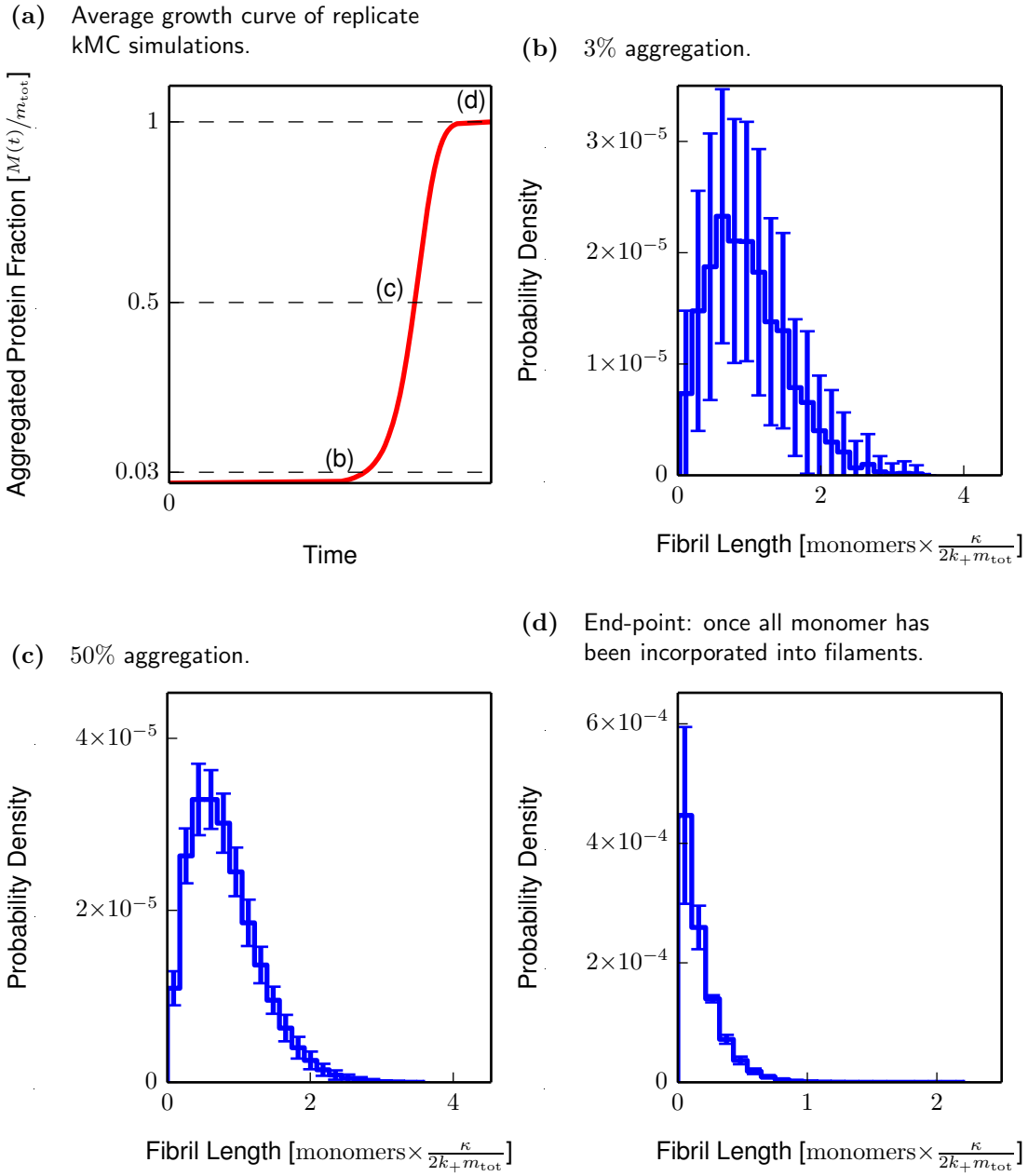


Figure 3.9 Average filament length distributions **(b–d)** at specific points on the average growth curve **(a)** of replicate kMC simulations of the fragmentation model, where the parameter values were $m_{\mathrm{tot}} = 10\ \mu\mathrm{M}$, $n_{\mathrm{c}} = 2$, $N(0) = 0$, $k_{\mathrm{n}} = 3\times10^{-7}\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$, $k_+ = 5\times10^{4}\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$, $k_- = 0$, $k_{\mathrm{f}} = 3\times10^{-8}\ \mathrm{s}^{-1}$, and $V = 83\times10^{-14}\ \mathrm{l}$. Filament lengths have been normalised by the predicted mean filament length $2k_+m_{\mathrm{tot}}/\kappa$ (equation 1.25 with the appropriate κ for the model). Each distribution represents the mean of 150 replicate simulations with error bars indicating the standard deviation (determined using the method in §2.2).

times. This intermediate distribution emerges once the monomer population has become depleted and the length distribution begins collapsing into the long-time steady state. At this point, longer filaments break and are not replaced. This leads to a rapid increase in the number of short filaments, producing the exponential distribution seen in figure 3.9d. The shape of the early-time and end-point distributions shown here will be important in §6 for interpreting experimentally determined filament length distributions.

It should be noted that, in a histogram where the bins are for ranges of filament lengths as opposed to individual filament lengths, it would be difficult to distinguish between this exponential and the very narrow peaked distribution of the long-time steady state, due to coarse-graining. However, the former is notably broader than the latter.

The error bars on each distribution in figure 3.9 indicate that the variability between replicate simulations is highest at early times and in the number of short filaments in the end-point distribution. The higher variance at early times is simply a result of fewer filaments being present, making fluctuations more apparent. Similarly, the large variance in the number of short filaments present in figure 3.9d is caused by variations in the time at which the last few monomers are incorporated into filaments and hence the differing amount of time that fragmentation has had to generate short filaments.

In a similar way, the variability at early times is evident for the heterogeneous nucleation model in figure 3.10, although the variability is more uniform over the distribution than it was for the fragmentation model. However, the end-point distribution in figure 3.10d is clearly different from that produced by the fragmentation model and is caused by the same monomer depletion effect as was seen to affect the maximum aggregation rate in §3.3.2. At early times a Poisson distribution emerges as predicted above. This persists while sufficient free monomer is available for the creation of new filaments, but as it decreases fewer filaments are formed. At later times, the decline in formation of new filaments results in a peaked filament length distribution. This peak is expected to disappear once distribution has relaxed under the action of depolymerisation.

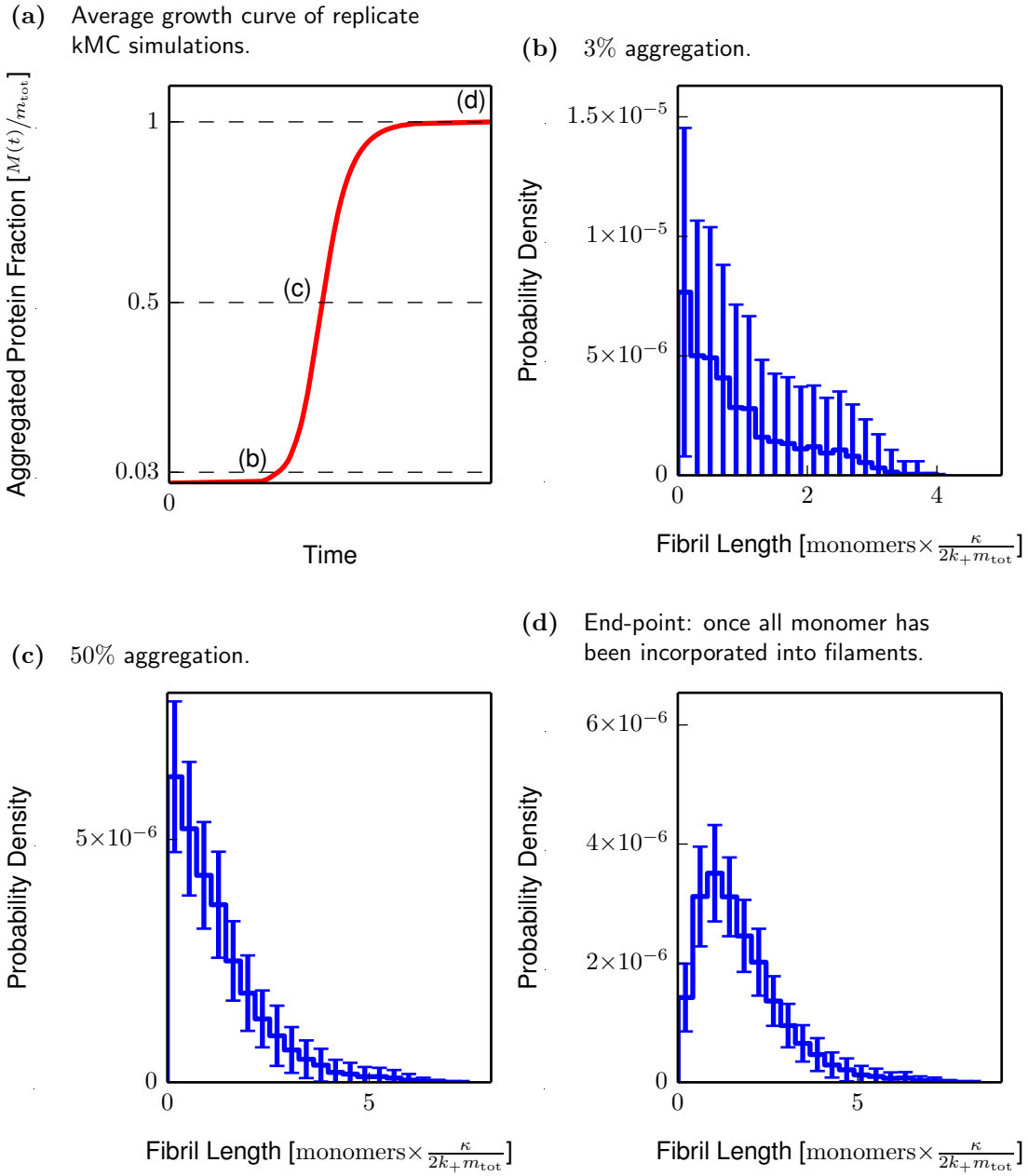


Figure 3.10 Average filament length distributions **(b–d)** at specific points on the average growth curve **(a)** of replicate kMC simulations of the heterogeneous nucleation model, where the parameter values were $m_{\mathrm{tot}} = 10\ \mu\mathrm{M}$, $N(0) = 0$, $n_{\mathrm{c}} = 2$, $k_{\mathrm{n}} = 3\times10^{-7}\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$, $k_+ = 5\times10^{4}\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$, $k_- = 3\times10^{-8}\ \mathrm{s}^{-1}$, $k_{\mathrm{h}} = 24\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$, $n_{\mathrm{s}} = 2$, and $V = 83\times10^{-14}\ \mathrm{l}$. Filament lengths have been normalised by the predicted mean filament length $2k_+m_{\mathrm{tot}}/\kappa$ (equation 1.25 with the appropriate κ for the model). Each distribution represents the mean of 150 replicate simulations with error bars indicating the standard deviation (determined using the method in §2.2).

Thus, stochasticity does not have a significant effect on the distribution of filament lengths, and the variability that is present in these distributions is due to fluctuations when a small number of filaments is present. However, for both models of autocatalysis I have identified an interesting intermediate length distribution which could not have been predicted from the analytic solution.

3.4 Discussion

Both the mean-field analytic solution and kinetic Monte Carlo simulations have their advantages and limitations when trying to understand the effects of various mechanisms on the formation of amyloid fibrils. The mean-field description allows the rate constants to be determined from experimental data, and provides predictions of the steady states which the system is expected to adopt at various points in the self-assembly process under the influence of particular mechanisms, both of which I shall be using again in §7. Stochastic simulations, on the other hand, are able to capture the effects of fluctuations which arise when the number of filaments present in the system is small, thereby highlighting the importance of rare events in individual mechanisms at specific points during the aggregation process.

The effects of stochasticity were most pronounced in the lag phase where the growth competent species are both short and few in number meaning fluctuations in either length or population have a large impact. The major contribution to this variability was the primary nucleation process, which also had an affect on the scaling behaviour of the lag-time with total protein concentration. By disabling this process and introducing a small concentration of preformed fibril seed I showed that the average behaviour of the replicate growth curves better matched the mean-field prediction; however, as I pointed out, some variability is still produced by autocatalytic polymerisation. Hence, the lag-time distribution, and the scaling of the lag time with total protein concentration, both need further investigation; this, I will present in §4 and §5, respectively.

A similar departure of the kinetic Monte Carlo simulation results, from the mean-field prediction, was also seen for the maximum aggregation rate. While the scaling trend was broadly reproduced it was apparent that at higher protein concentrations the aggregation rate was not as high as expected. This, I stated, was likely to be due to a lower number of filaments being present than theoretically possible, thereby restricting the rate at which monomer could be incorporated into filaments. If indeed caused by this, then the larger difference between the experimental results and the prediction would indicate that even fewer filaments are present. This observation, in conjunction with other experimentally measured differences between high and low concentrations of bovine insulin, lead to the formulation of a modified fragmentation model in which fragmentation is arrested and thereby limits the number of filaments. The effects of this model form the subject of §6.

Chapter 4

The effects of primary nucleation and autocatalysis on the lag-time distribution

4.1 Introduction

Primary nucleation is thought to be an inherently stochastic process and that its fluctuations should be detectable as variability in the lag time of replicate experiments [45, 48, 111]. Hence, accurate measurement of the lag-time variability under different conditions could allow the primary nucleation process to be probed and may also provide a way to identify other processes important during the lag phase. However, care is needed in interpreting the experimental data: it has been noted that the lag-time variability can be markedly reduced by careful selection of the environmental conditions [74].

A distribution of τ_{ϕ_M} lag times for bovine insulin (from the growth curves in figure 3.1) is shown in figure 4.1 for a threshold fraction of $\phi_M = 0.1$. Similar distributions are obtained for other protein concentrations, and values of ϕ_M in the range $0.03 \lesssim \phi_M \leq 0.1$.

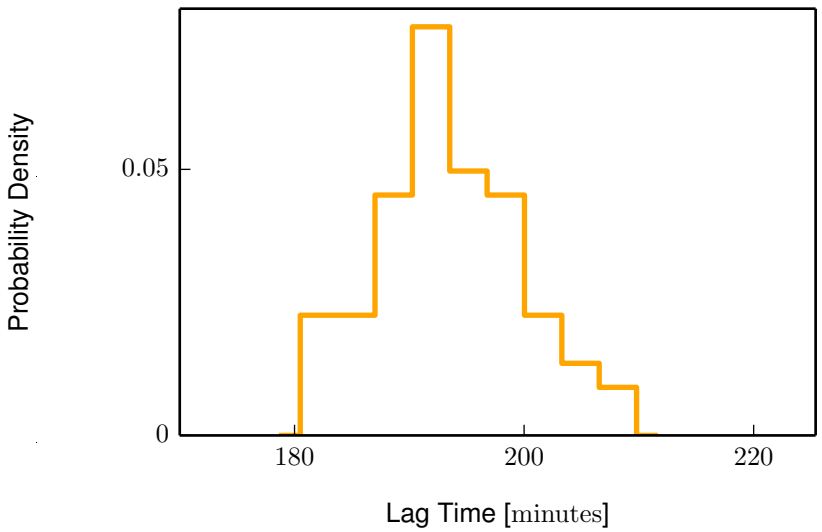


Figure 4.1 The distribution of lag times from the growth curves of 131 μM bovine insulin (shown in figure 3.1), for the lag time defined as the point the growth curve when 10% of the maximum fluorescence value is reached. Similar shape distributions are found for other concentrations of bovine insulin, and when the lag times are extracted by fitting the maximum aggregation slope (see §2.3.2).

As I noted in §3, the kinetic Monte Carlo simulations displayed more variability in the length of the lag phase than in any other aspect of the growth curve. While this variability is substantially larger when primary nucleation is active (i.e. $k_n > 0$), considerable variability is still produced by autocatalysis. This is apparent in figure 4.2, which compares the distribution of lag times produced by the fragmentation model, both with and without primary nucleation.

When primary nucleation is active, the distribution of lag times becomes close to exponential, while those produced by autocatalysis (in the absence of primary nucleation) have a much narrower, peaked distribution. In this chapter, I will show that the exponential-like distribution arises when primary nucleation is treated as a single rare event (i.e. formation of the first primary nucleus), while the peaked distribution emerges from the fact that the time at which the threshold is reached depends on a sequence of processes with random waiting-times such that — when the number of steps in this sequence is large — the distribution of times to reach the threshold can be described by central limit theorem [89].

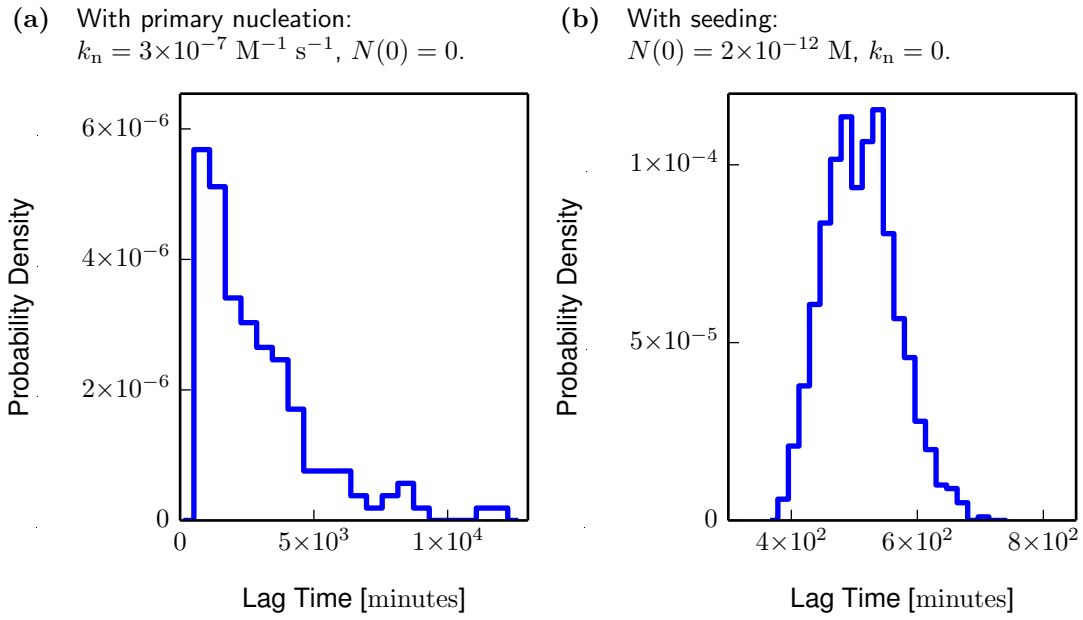


Figure 4.2 The distribution of lag times from kMC simulation growth curves of the fragmentation model (§3) when: **(a)** primary nucleation is active, as in figure 3.4; or **(b)** self-assembly is seeded, as in figure 3.6b. The common parameter values were $m_{\text{tot}} = 10$ µM, $n_c = 2$, $k_+ = 5\times10^4$ M^{-1} s^{-1}, $k_- = 0$, $k_f = 3\times10^{-8}$ s^{-1}, and $V = 83\times10^{-14}$ l. Here, the lag time is defined as the point in the growth curve when 10% of the total protein concentration had been incorporated into filaments. Similar shape distributions are found for other values of m_{tot}, and when the lag times are extracted by fitting the maximum aggregation slop (see §2.3.2).

Interestingly, the simulation lag-time distribution in figure 4.2b is similar in appearance to the experimental distribution in figure 4.1; this is unexpected as the experimental kinetics are widely thought to involve primary nucleation. This observation suggests that fluctuations within the autocatalytic process may be responsible for a significant proportion of the variability seen experimentally.

The time at which a threshold concentration of aggregated protein $M_\phi = \phi_M\, m_{\text{tot}}$ is reached, can be thought of as a first passage problem. First passage problems are widely studied in statistical physics in situations where one wishes to compute the probability distribution for the time at which a stochastic process first reaches a defined threshold. In some cases these can be solved analytically [89].

By considering autocatalytic polymerisation as a stochastic first passage problem, I will explore the waiting-time distributions associated with a generic autocat-

alytic feedback process in §4.3. First however, I will discuss the waiting-time distributions for sequential polymerisation-like processes, as these are relevant to the early kinetics as well as the discrete formation of a nucleus when nucleation is not treated simply as a single step process. To do this, I will use "minimal models" that focus on specific mechanisms or combinations of mechanisms. While these explicitly neglect the contribution of all other mechanisms thought to be involved in amyloid fibril self-assembly, the simplification means that the models can (in most cases) be solved analytically with fewer assumptions. From these results, complex behaviour can be understood in terms of very simple interactions, thereby providing a way to identify the key contributors to specific aspects of the early-time kinetics.

4.2 Waiting-time distributions for sequential polymerisation-like processes

In this section I will model the self-assembly of fibrils as a random walker moving along a line, where rightward steps correspond to an increase in fibrillar protein content, and leftward steps correspond to a decrease in fibrillar protein content. The aggregate being described may be a single filament, or an ensemble where I consider the combined total monomer content of all filaments and not the individuals.

In the case where multiple aggregates compete, multiple random walkers will be considered, each describing the length of an individual filament. The first of these $\mathcal{N}_\mathrm{s}$ competing aggregates to reach the threshold size will thus define the first passage time. As an example, it is believed that in primary nucleation, multiple aggregates fluctuate in size through reversible, polymerisation-like processes until one reaches a size n_c. Thus, some of the steps along the line may have different rates; for example, when the rates are time dependent due to effects such as monomer depletion, or in the case of an energy barrier to nucleation.

Where possible, I will use the same parameter values as in other chapters, but it should be noted that many of the example distributions given below tend toward the central limit theorem prediction when the number of steps involved, the monomer concentrations, or the rate constants become large. For computational convenience (reduced run-time) I will use a small threshold, $\phi_M = 0.03$.

A small value of ϕ_M also means that the monomer population at the first passage time will be close to its starting value, and so I can neglect monomer consumption during the derivation of the analytic expressions for waiting-time distributions throughout this section; thus, most of the analytic expressions that I present here can be expected to only give accurate results for waiting-time distributions when $\phi_M \lesssim 0.1$.

4.2.1 Waiting times to reach an aggregation threshold by reversible polymerisation

First, I will focus on the distribution of times taken to reach the threshold quantity of aggregated protein M_ϕ when started from a preformed nucleus of size n_c where the only processes involved are polymerisation and depolymersiation. A first passage time problem of this kind can be solved analytically, for the moments of the probability distribution, using the one-dimensional discrete Fokker-Plank equation [28], which is used to describe problems where a single particle is moving on a one-dimensional lattice under the actions of diffusion and drift.

The one-dimensional lattice has two boundary conditions: a reflecting boundary at $a = n_\mathrm{c}$, which prevents the possibility of the fibril disappearing; and an absorbing boundary at $b = M_\phi N_\mathrm{A} V$, which is the end point where aggregation has reached the threshold. The position of the walker, representing the length of the filament, can be changed by both addition and loss of monomers. This gives a diffusion constant $D = \lambda_+ + \lambda_-$, where $\lambda_+ = 2k_+ m_\mathrm{tot} \approx 2k_+ m(t)$ is the elongation propensity and $\lambda_- = 2k_-$ is the depolymerisation propensity. Since filaments on average tend to grow over time, it will be assumed that $\lambda_+ \gg \lambda_-$. This bias toward growth can be described as drift, with rate $\omega = \lambda_+ - \lambda_-$. Thus, the walker describing the fibril's length starts at lattice site a and performs a biased random walk until it reaches b.

Using the approach outlined in §A1 [28], the first two moments of the first passage probability distribution were found to be

$$T_1(x) = \mu = \frac{1}{\omega}\left(b - x + \frac{D}{2\omega}\left(e^{-\frac{2\omega}{D}(b-a)} - e^{-\frac{2\omega}{D}(x-a)}\right)\right) \tag{4.1}$$

and

$$\begin{aligned} T_2(x) = (\sigma^2 + \mu^2) =& \frac{(x-b)^2}{\omega^2} \\ &+ \frac{D}{\omega^3}\Bigg[\quad 2a\left(e^{-\frac{2\omega}{D}(x-a)} - e^{-\frac{2\omega}{D}(b-a)}\right) \\ &\qquad +b\left(1 + 3e^{-\frac{2\omega}{D}(b-a)} - e^{-\frac{2\omega}{D}(x-a)}\right) \\ &\qquad -x\left(1 + e^{-\frac{2\omega}{D}(x-a)} - e^{-\frac{2\omega}{D}(b-a)}\right)\Bigg] \\ &+ \frac{D^2}{\omega^4}\Bigg[\quad e^{-\frac{2\omega}{D}(b-a)} - e^{-\frac{2\omega}{D}(x-a)} \\ &\qquad +\frac{1}{2}e^{-\frac{4\omega}{D}(b-a)} - \frac{1}{2}e^{-\frac{2\omega}{D}(x+b-2a)}\Bigg] \end{aligned} \tag{4.2}$$

from which the mean $\mu = T_1(x)$ and standard deviation $\sigma = \sqrt{T_2(x) - T_1(x)^2}$ may be obtained.

As the shape of the distribution cannot be calculated analytically, I will assume that the waiting times follow a Gaussian distribution defined by these two moments; this Gaussian approximation is shown in figure 4.3 together with kinetic Monte Carlo simulation results for reversible polymerisation, for different rates of depolymerisation. For the parameters used, this approximation appears reasonable, producing very similar shaped distributions to the simulation results.

The increased distribution width when the depolymerisation rate is higher is the result of an increase in the number, and the variability in the number, of steps (both polymerisation and depolymerisation) necessary to reach the threshold. Similarly, the shift of the distribution, indicating a longer mean waiting-time, is also the result of the increased number of steps when the depolymerisation rate is higher. Both of these can be understood by examining how the diffusion

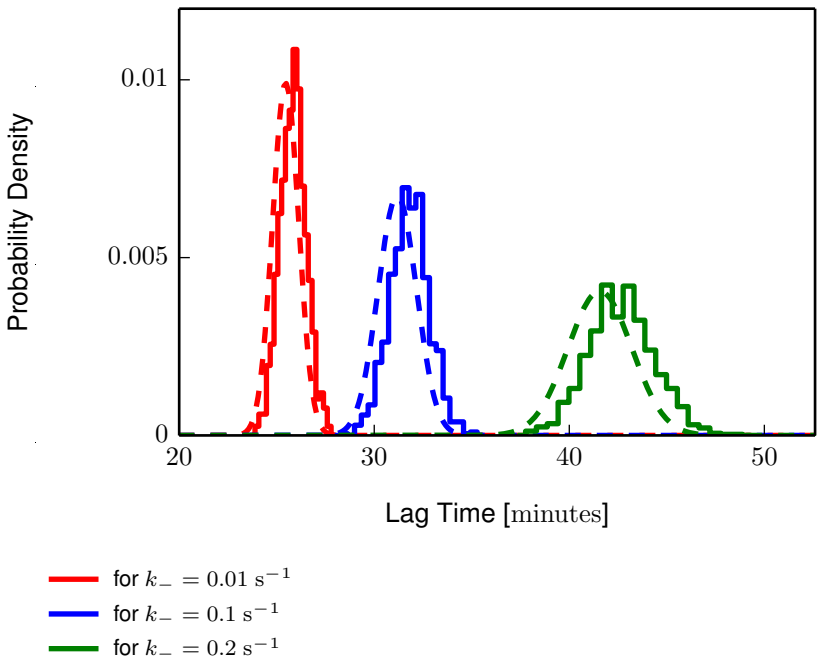


Figure 4.3 Theoretical probability distributions of waiting times for a single filament to reach a threshold length by reversible polymerisation, for different depolymerisation rates, k_-; the common parameter values were $m_{\mathrm{tot}} = 10\ \mu\mathrm{M}$, $n_{\mathrm{c}} = 2$, $k_{\mathrm{n}} = 0$, $k_+ = 5{\times}10^4\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$, and $V = 83{\times}10^{-16}\ \mathrm{l}$. The **histograms** contain waiting times from 1000 replicate kMC simulations, and the **dashed lines** are Gaussian approximations for the theoretical predictions of the mean and variance, given by equations 4.1 and 4.2, respectively.

constant D and drift constant ω change in opposition to one another with the depolymerisation rate.

The slight offset of the mean of the theoretical prediction with respect to the simulation data is a consequence of monomer depletion, which has been neglected in deriving equations 4.1 and 4.2 but does affect the simulations. Hence, for a higher depolymerisation rate and increased number (and variability in the number) of steps, this offset becomes larger.

Depolymerisation is often neglected when dealing with amyloid fibril formation [55, 81, 82]. This is typically because the rate of depolymerisation is vanishingly small compared to other processes until the number of filament ends becomes large and the monomer population has been depleted; conditions associated with later times. At early times however, the monomer population will

be high, and elongation and primary nucleation will be much more significant than depolymerisation. Therefore, for the rest of this chapter I will neglect depolymerisation, and instead focus on elongation and nucleation events which I will model using random walks with no backward stepping.

4.2.2 The Erlang distribution: waiting times to reach the k^{th} Poisson event

The waiting-time distribution for a sequence of k Poisson events, each with individual rate λ, is given by the Erlang distribution

$$P(\tau; k, \lambda) = \frac{\lambda(\lambda\tau)^{k-1}e^{-\lambda\tau}}{(k-1)!} \tag{4.3}$$

The mean and variance of this peaked distribution are defined through the relations $^{k}/_{\lambda}$ and $^{k}/_{\lambda^2}$, respectively [89].

Below, I will show how this is useful for modelling rapid primary nucleation and later for polymerisation with and without slow primary nucleation (§4.2.4).

No nucleation barrier

If primary nucleation is rapid (i.e. there is no appreciable nucleation barrier) and treated as a single-step process (i.e. $k_{\text{n}} \approx k_{+}$), then the number of growth-competent ends will be very high. This will reduce the free monomer population available for elongation and result in a large number of short filaments. Consequently, the major contribution to reaching M_{ϕ} will come from nucleation, and not elongation which can be neglected. This is not a particularly realistic picture for fibril formation but I have included it for the sake of completeness.

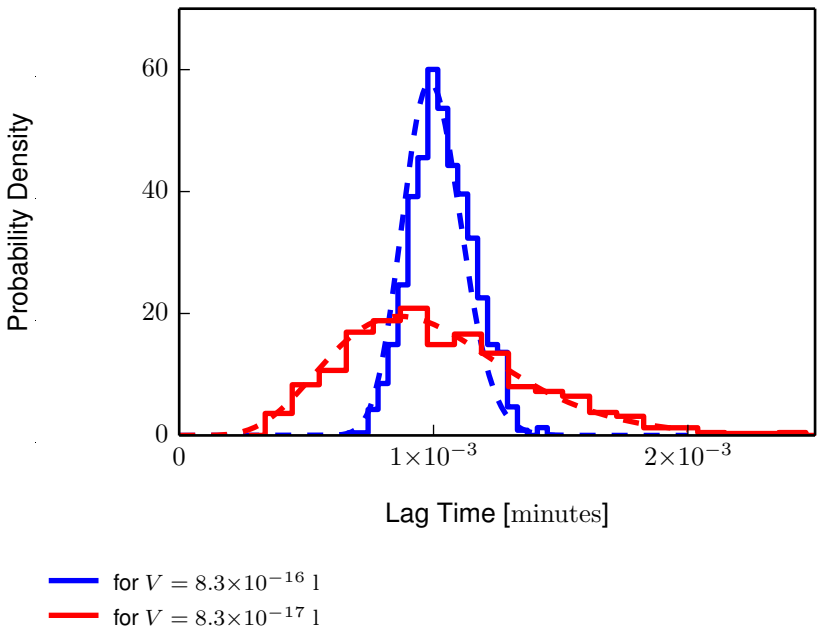


Figure 4.4 Theoretical probability distributions of waiting times for rapid primary nucleation to incorporate a threshold fraction of the total protein concentration into filaments, for different system volumes, V; the common parameter values were $m_{\mathrm{tot}} = 10$ μM, $n_{\mathrm{c}} = 2$, $k_{\mathrm{n}} = 5{\times}10^{4}$ M^{-1} s^{-1}, $k_{+} = 0$, and $k_{-} = 0$. The **histograms** contain waiting times from 1000 replicate kMC simulations, and the **dashed lines** are the theoretical prediction by equation 4.3.

Assuming that only primary nucleation contributes toward M_{ϕ}, then the number of nuclei required is $k = {}^{N_{\mathrm{A}} V M_{\phi}}/_{n_{\mathrm{c}}}$, with rate $\lambda = {}^{\eta}/_{N(0)} = {}^{N_{\mathrm{A}} V k_{\mathrm{n}} m_{\mathrm{tot}}^{n_{\mathrm{c}}}}/_{n_{\mathrm{c}}!}$, where ${}^{N(0)}/_{\eta}$ is the rate of primary nucleation given in §3.3.3 assuming a constant monomer concentration $m(t) \approx m_{\mathrm{tot}}$. Figure 4.4 shows the distributions obtained for waiting times derived from both equation 4.3 and kinetic Monte Carlo simulations, for different values of V. The widths of these distributions depend inversely on both V and m_{tot} because fluctuations in the number of steps involved are greater for a smaller number of particles, while the mean only depends on m_{tot}. The slight discrepancy between the predicted and simulated distributions is due to monomer depletion effects, but the data are otherwise well described by the Erlang distribution.

4.2.3 Competition between aggregates: the formation of the first primary nucleus

I now consider the case where primary nucleation is very slow. Here, the formation of the first primary nucleus can be described as a competition between multiple growing aggregates, each smaller than the nucleus size n_c. The first of these to reach the critical size n_c will then typically dominate the subsequent growth. I will treat this competition as a number of parallel one-dimensional random walks, each with an Erlang distribution of first passage times.

In order to obtain a closed analytic expression for the distribution of waiting times for the first of these Erlang processes to reach M_ϕ, the number of competing processes $\mathcal{N}_s$, must be known in advance. From a simulation point of view, $\mathcal{N}_s$ can be fixed by defining "special" monomers which are known in advance to initiate the formation of a nucleus, even though the time at which each does so, is not known. It will be assumed that the $\mathcal{N}_s$ "special" monomers are indistinguishable from any other monomer in all ways except that they cannot join to one another; hence, each will only contribute to $M(0)$ once it has formed an aggregate of the minimal stable size n_c (so that $m(0) = m_{tot}$).

Initially I will treat these $\mathcal{N}_s$ "special" monomers as completely non-interacting. This is not strictly true since they compete with one another through the monomer population, but that case is a modification of the following.

A non-interacting race

Given $\mathcal{N}_s$ non-interacting elongation-like processes of the kind defined above, the probability distribution for each to be the first to reach the critical length M_ϕ at time τ is given by

$$P(\tau; k, \lambda, \mathcal{N}_s) = F(\tau; k, \lambda) \left(1 - \int_0^\tau F(t; k, \lambda) dt\right)^{\mathcal{N}_s - 1} \tag{4.4}$$

where $F(t;k,\lambda)$ is the Erlang probability distribution given by equation 4.3, and the parameters k and λ have the same meaning as in §4.2.4. The distribution of waiting times for the first out of all $\mathcal{N}_s$ aggregates to reach the threshold, can thus be obtained by multiplying equation 4.4 by $\mathcal{N}_s$:

$$P\left(\tau;k,\lambda(\tau),\mathcal{N}_s\right)=\mathcal{N}_s\frac{\lambda(\tau)(\lambda\tau)^{k-1}e^{-\lambda\tau}}{\Gamma(k-1)}\left(\frac{\Gamma(k,\lambda\tau)}{\Gamma(k)}\right)^{\mathcal{N}_s-1} \tag{4.5}$$

Although equation 4.5 describes the outcome of a competition to form the first primary nucleus, I will deliberately define the number of steps $k=M_\phi N_A V-n_c$ in terms of n_c [(1)] and M_ϕ [(2)], and the rate $\lambda=2k_+m_{tot}\approx 2k_+m(t)$ at which these steps occur in terms of the elongation rate k_+, so as to illustrate the distributions produced by this model in a way that can be directly compared with the other results in this chapter.

Equation 4.5 and the distribution of waiting times obtained from kinetic Monte Carlo simulations are shown in figure 4.5 for different system sizes and values of $\mathcal{N}_s$. When $\mathcal{N}_s$ is small, the number of steps required to reach the threshold is much smaller than the number of monomers per aggregate ($k\ll {}^{m_{tot}}/_{N(0)}$), and equation 4.5 agrees well with simulation. However, as $\mathcal{N}_s$ is increased, the fibrils interact via the free monomer pool, causing the approximation $m(t)\approx m_{tot}$ to break down such that agreement between equation 4.5 completely fails once $k\approx {}^{m_{tot}}/_{N(0)}$. Fortunately it is simple to account for the rate of monomer consumption, as I will now demonstrate.

Time dependent propensity: accounting for monomer depletion

The time dependence of $m(t)$ can be accounted for by re-deriving the Erlang distribution with $\lambda=\lambda(t)$, and then reintegrating equation 4.4 (shown in §A2). This introduces the integral $\Lambda(\tau)=\int_0^\tau\lambda(t)dt$ in place of $\lambda\tau$ in equation 4.3. The waiting-time distribution for each of the $\mathcal{N}_s$ competing aggregates is then given by

$$P(\tau;k,\lambda(\tau))=\frac{\lambda(\tau)(\Lambda(\tau))^{k-1}e^{-\Lambda(\tau)}}{(k-1)!} \tag{4.6}$$

(1) The starting size of each of the $\mathcal{N}_s$ aggregates in this example is larger than a single monomer; hence these contribute to the aggregated monomer count.

(2) Rather than a fixed aggregate size, this is a fraction of the total protein concentration.

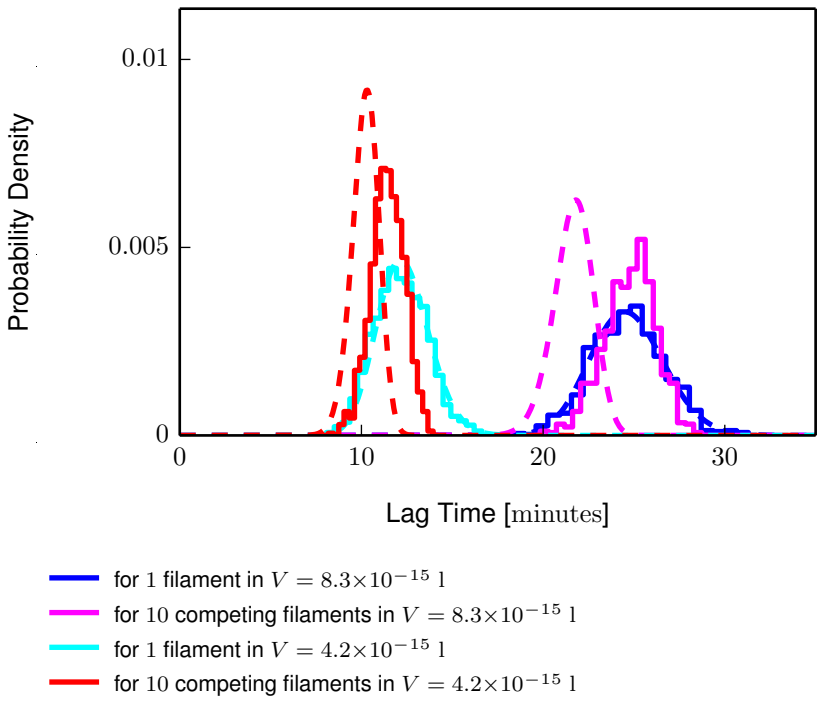


Figure 4.5 Theoretical probability distributions of waiting times for the first of $\mathcal{N}_s$ competing filaments to incorporate a threshold fraction of the total protein concentration into itself, for different numbers of competing filaments, and system volumes, V; the common parameter values were $m_{tot} = 1$ μM, $n_c = 2$, $k_n = 0$, $k_+ = 5\times10^4$ M^{-1} s^{-1} and $k_- = 0$. The **histograms** contain waiting times from 1000 replicate kMC simulations, and the **dashed lines** are the theoretical prediction by equation 4.5.

from which the waiting-time distribution for the first aggregate to reach the critical size can be obtained:

$$P\left(\tau; k, \lambda(\tau), \mathcal{N}_s\right) = \mathcal{N}_s \frac{\lambda(\tau)(\Lambda(\tau))^{k-1} e^{-\Lambda(\tau)}}{\Gamma(k-1)} \left(\frac{\Gamma(k, \Lambda(\tau))}{\Gamma(k)}\right)^{\mathcal{N}_s - 1} \tag{4.7}$$

The rate of change of monomer concentration for a system involving only polymerisation is

$$\frac{dm(t)}{dt} = -2N(t)k_+m(t)$$

Since the number of growth competent aggregates $N(t)$ is constant in this model [3] and the monomer concentration is conversed ($m_{tot} = m(t) + M(t)$),

[3] In the absence of autocatalysis and primary nucleaiton, this is fixed by the number of "special" monomers.

this equation can be solved exactly and the parameter $\lambda(\tau)$ obtained from $m(t)$, giving

$$\lambda(\tau) = 2k_+ m_{\text{tot}} e^{-2k_+ N(0)\tau}$$

from which

$$\Lambda(\tau) = \frac{m_{\text{tot}}}{N(0)} \left(1 - e^{-2k_+ N(0)\tau}\right)$$

where $N(0) = {}^{\mathcal{N}_\text{s}}/_{N_\text{A} V}$ and N_A is Avogadro's constant. Again I have assumed that the $\mathcal{N}_\text{s}$ "special" monomers that are present at the start do not contribute to $M(0)$ (so that $m(0) = m_{\text{tot}}$). However, the assumption that $m_{\text{tot}} \gg N(0)$, on which this is based, is certainly valid because primary nucleation has already been defined as a rare occurrence.

As demonstrated in figure 4.6, equation 4.7 is in much better agreement with the simulation data than was equation 4.5. However, equation 4.7 still fails when the number of steps required is greater than the average amount of monomer per aggregate, $k \gtrsim \frac{m_{\text{tot}}}{N(0)}$. This condition is outwith the scope of this solution because they assume a higher rate of primary nucleation, producing a larger number of competing nuclei, in which case there would be a non-zero probability that none of the fibrils reach M_ϕ because competition is too high and the total monomer population is too small to be shared among the number of aggregates; i.e. the monomer pool is used up in the subcritical aggregates. That said, fluctuations in length caused by depolymerisation could eventually lead to one of them reaching the threshold, but the associated waiting-time distribution for such events is expected to be very broad due to the number of steps involved.

This example is also not strictly for primary nucleation but is a type of first-past-the-post process. If equation 4.7 were used to describe primary nucleation more realistically, k would be a constant, independent of V and m_{tot} (something like $k = n_\text{c} - 1$, for example). Conversely, λ would still depend on the number of competing proto-nuclei and hence V, and thus the variance in the waiting-time distribution can be expected to scale far more strongly with the system size than is evident in figure 4.6; i.e. the distribution would reach the deterministic limit more rapidly than is apparent in the figure.

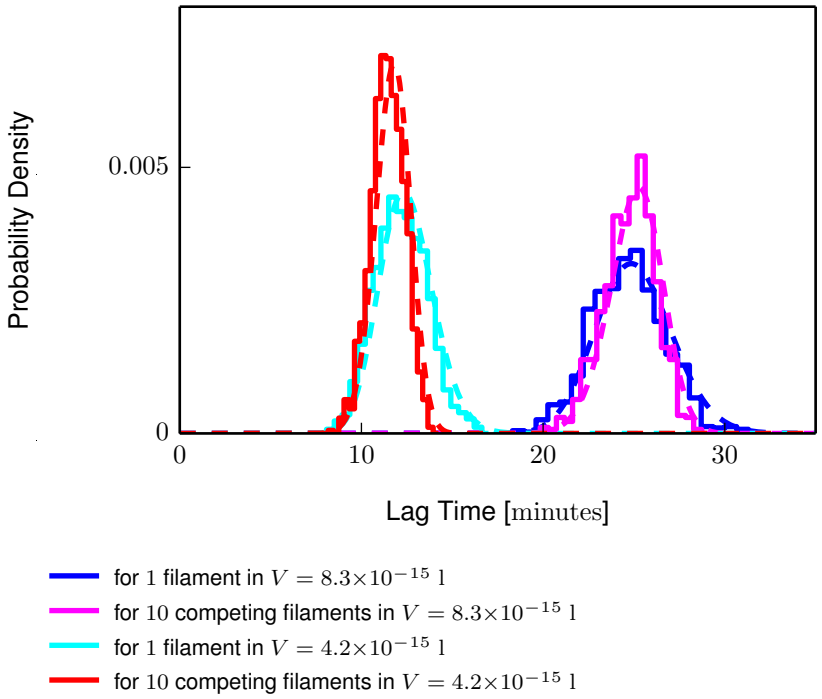


Figure 4.6 Theoretical probability distributions of waiting times for the first of $\mathcal{N}_s$ competing filaments to incorporate a threshold fraction of the total protein concentration into itself with a time dependent rate, for the same parameter values as used in figure 4.5. The **histograms** contain waiting times from 1000 replicate kMC simulations and are identical to those in figure 4.5, and the **dashed lines** are the theoretical prediction by equation 4.7, where $\Lambda(\tau) = \frac{m_{\text{tot}}}{N(0)}\left(1 - e^{-2k_+ N(0)\tau}\right)$.

4.2.4 Waiting times for the combined contribution from multiple polymerisation processes

In the case of amyloid fibril formation, it is commonly thought that the elongation of a number of filaments collectively contribute to reaching the threshold M_ϕ. These could be formed through primary nucleation or autocatalysis, but for simplicity I will first consider seeded self-assembly where a fixed number of filaments $\mathcal{N}_s = N(0)N_A V$ is present throughout, and the only active process is filament elongation.

The waiting times for this elongation model are again be described by the Erlang distribution (in §4.2.2), where the number of steps required for these seeds to

collectively reach the threshold is $k = M_\phi N_\mathrm{A} V - \mathcal{N}_\mathrm{s} n_\mathrm{c}$ [4], and the stepping rate $\lambda = 2\mathcal{N}_\mathrm{s} k_+ m_\mathrm{tot} \approx 2\mathcal{N}_\mathrm{s} k_+ m(t)$ [5], is given by the propensity in reaction 1.8 for $\mathcal{N}_\mathrm{s}$ simultaneously growing aggregates.

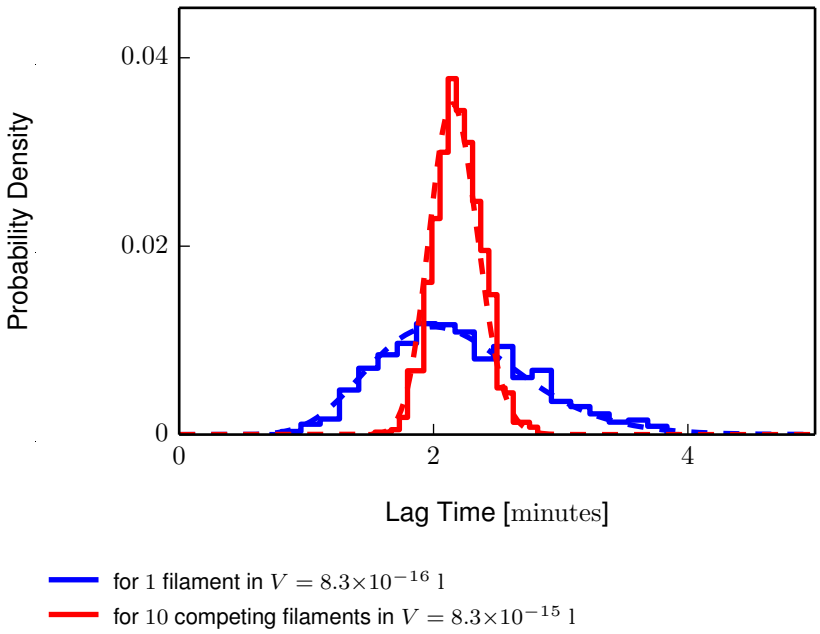


Figure 4.7 Theoretical probability distributions of waiting times for $\mathcal{N}_\mathrm{s}$ filaments to accrete a threshold fraction of the total protein concentration, collectively, for different numbers filaments, and system volumes, V; the common parameter values were $m_\mathrm{tot} = 1$ μM, $n_\mathrm{c} = 2$, $k_\mathrm{n} = 0$, $k_+ = 5\times10^4$ M^{-1} s^{-1} and $k_- = 0$. The **histograms** contain waiting times from 1000 replicate kMC simulations, and the **dashed lines** are the theoretical prediction by equation 4.3.

Figure 4.7 shows waiting-time distributions for two different system volumes V, containing the same concentration of seeds $N(0)$. These distributions were obtained from kinetic Monte Carlo simulations only involving monomer addition, and the theoretical prediction from equation 4.3. From the definitions in §4.2.2, the width of these distributions are determined by V and m_tot, and the mean by m_tot. However, $\mathcal{N}_\mathrm{s}$ also plays an important role in both the mean and variance because it affects the amount of material already present at $t = 0$ and the rate at which new material is polymerised. Thus, although the concentration of seeds is the same, the waiting-time distribution in a larger volume becomes narrower because the higher rate of accretion of monomers produces a shorter waiting-time

[4] Here, the seeds contribute towards the threshold.
[5] Assuming an approximately constant monomer concentration at early times.

between steps, and fewer monomers per seed filament are needed in order for the threshold to be reached.

The distributions in figure 4.7 are expected in the case where polymerisation is the sole contribution to the lag time. However, primary nucleation was seen to dramatically alter the lag-time distribution in the kinetic Monte Carlo simulations (figure 4.2). It is therefore also important to consider the effects of primary nucleation on the lag-time distribution.

A single rate-limiting step: the effects of primary nucleation

When incorporating primary nucleation into the polymerisation model presented above, it is necessary to treat the process as a single slow step; in difference to previously treating it as a sequence of steps. This is because the first passage time for the sequence of steps only has a closed form solution when at most one step in the sequence has a different rate. For this slow step, I will use the same propensity as given in §4.2.2 (i.e. $\alpha = \eta/N(0)$) and assume that the step produces $\mathcal{N}_\mathrm{s} = N(0)N_\mathrm{A}V = 1$ nucleus [6].

Using the approach described in §A2 — which is a variation of the derivation of the Erlang distribution used above — I obtain the waiting-time distribution for polymerisation:

$$P(\tau; k, \lambda, \alpha) = \frac{\alpha\lambda^k e^{-\alpha\tau}}{(\lambda-\alpha)^k}\left[1 - \frac{\Gamma(k, (\lambda-\alpha)\tau)}{\Gamma(k)}\right] \tag{4.8}$$

where a single slow step occurs somewhere (unspecified) in the sequence. The term $1 - \Gamma(k, (\lambda-\alpha)\tau)/\Gamma(k)$ in the brackets is the cumulative distribution function for the gamma distribution [7].

[6] While I only consider the case where a single nucleus is present, this expression is valid for $\mathcal{N}_\mathrm{s} \geq 1$.

[7] The gamma distribution is a generalised form of the Erlang for which k may be non-integer.

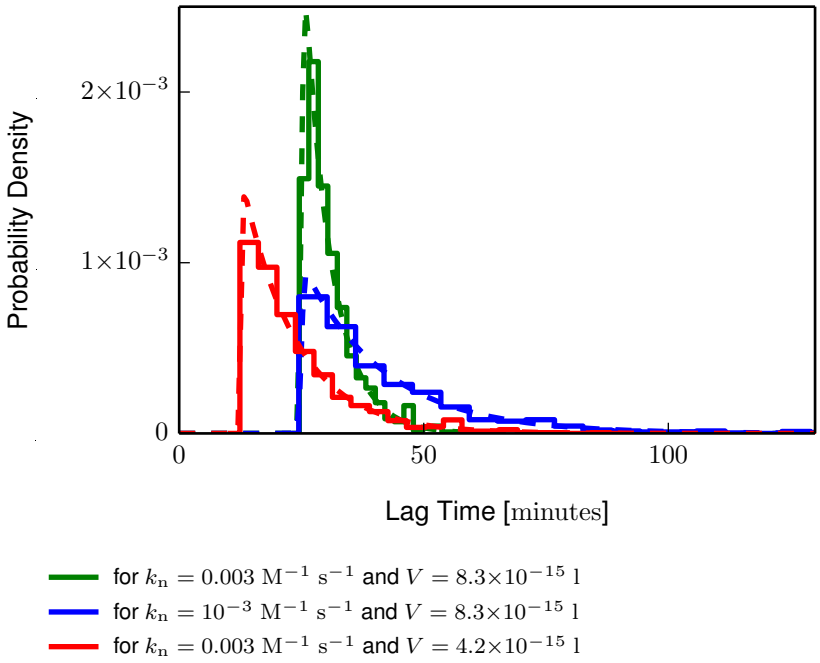


Figure 4.8 Theoretical probability distributions of waiting times for $\mathcal{N}_s$ filaments, created by primary nucleation, to accrete a threshold fraction of the total protein concentration, collectively, for different primary nucleation rates, k_n, and system volumes, V; the common parameter values were $m_{tot} = 1$ μM, $n_c = 2$, $k_+ = 5\times10^4$ M^{-1} s^{-1} and $k_- = 0$. The **histograms** contain waiting times from 1000 replicate kMC simulations, and the **dashed lines** are the theoretical prediction by equation 4.8.

When $\alpha \ll \lambda$, the slow step (representing primary nucleus formation) dominates the waiting-time distribution, resulting in an exponential distribution of lag times with an offset as illustrated in figure 4.8 for different nucleation rates. This distribution has the same shape as the waiting-time distribution of the slow step itself, indicating that the waiting time is being almost entirely determined by the single slow step. The offset is produced by the fact that there is a minimum time taken by the elongation steps in order to reach M_ϕ, once the primary nucleus has formed. For higher values of α, the distribution rapidly tends towards the shape seen in figure 4.7 (i.e. without a slow step). Thus, rare primary nucleation can dominate the lag time, but at higher primary nucleation rates the lag time due to this one step can become negligible compared to the large number of elongation steps. Both of these cases can be shown by series expansion of equation 4.8: the former in the limit $\alpha/\lambda \to 0$, the latter in the limit $\alpha \approx \lambda$.

4.2.5 Implications for the autocatalysis simulations

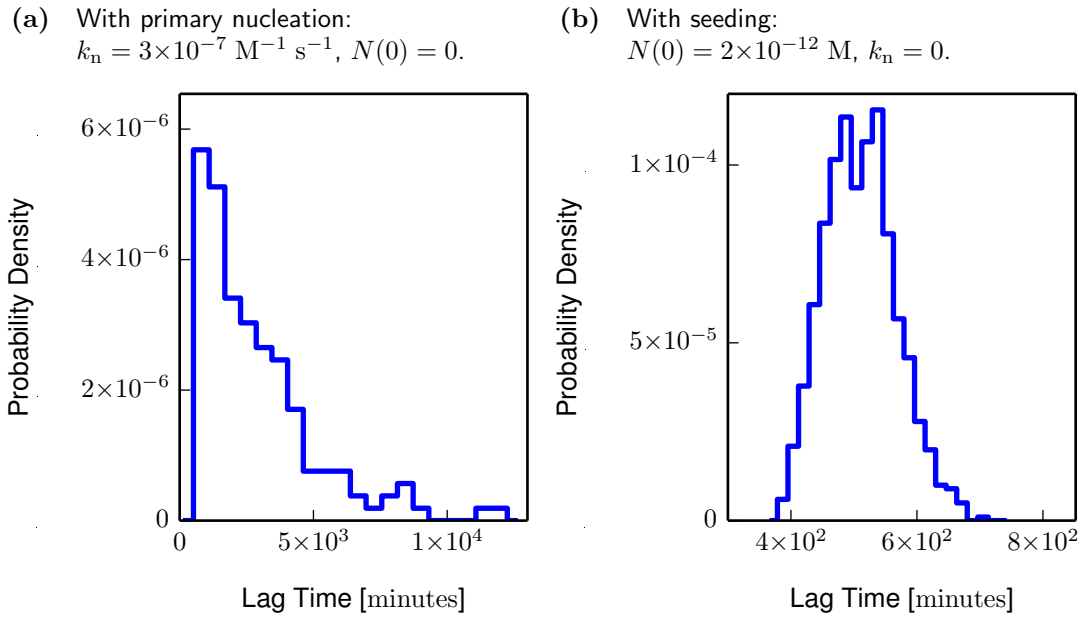


Figure 4.9 The distribution of lag times from kMC simulation growth curves of the fragmentation model (previously shown in figure 4.2) when: **(a)** primary nucleation is active, as in figure 4.2a; or **(b)** self-assembly is seeded, as in figure 4.2b.

The model for amyloid fibril formation that best describes bovine insulin requires primary nucleation, elongation and at least one of either fragmentation or heterogeneous nucleation. In the above models I have ignored both of the autocatalysis mechanisms and focused only on primary nucleation and elongation. Hence, the analytic expressions that I derived are not be able to quantitatively describe the distributions in figure 4.9, which were produced by kinetic Monte Carlo simulations involving filament fragmentation. However, even in these it is evident that primary nucleation can dominate the lag time and result in an exponential distribution similar to the one given above for polymerisation with a single slow step (figure 4.8).

4.3 Waiting-time distributions for autocatalytic polymerisation

Autocatalysis will accelerate aggregation by providing additional growth-competent filament ends. Compared to nucleation-dependent polymerisation, this will reduce the mean time to reach the aggregation threshold but it is not obvious what effect it will have on the waiting-time distribution. In order to obtain a closed form, analytic expression for the lag-time distribution produced by autocatalytic polymerisation, equations 1.12 and 1.13, which respectively describe the change in the number of filaments and concentration of fibrillar protein at time t, need to be greatly simplified. This can be achieved by considering a new "coarse-grained" model containing only two reactions — one increasing the quantity of protein in filaments in proportion to the current number of ends (due to elongation), and the other increase in the number of ends in proportion to the current quantity of protein in filaments (due to autocatalysis) — and comparing the closed form, analytic solution obtained for this model with kinetic Monte Carlo simulations of this simplified model and the original model.

4.3.1 The waiting-time distribution for a simplified autocatalytic polymerisation model

As in §4.2, I will neglect the effects of monomer depletion, which are expected to be small during the lag phase. With this assumption, the reactions describing elongation and the autocatalytic formation of new filaments can be simplified:

$$\mathcal{M} \xrightarrow{\mathcal{N}\mu} \mathcal{M} + 1 \tag{4.9}$$

$$\mathcal{N} \xrightarrow{\mathcal{M}\nu} \mathcal{N} + 1 \tag{4.10}$$

These describe the increase in the number of filaments $\mathcal{N}$ and their total monomer content $\mathcal{M}$ using the same feedback principle as contained in reaction 1.8 and 1.10, but with time independent general rate constants ν and μ, respectively [8].

[8] The rates ν and μ can be related to the fragmentation and elongation propensities thus: $\nu = k_{\mathrm{f}}$ and $\mu = 2m_{\mathrm{tot}}k_{+}$.

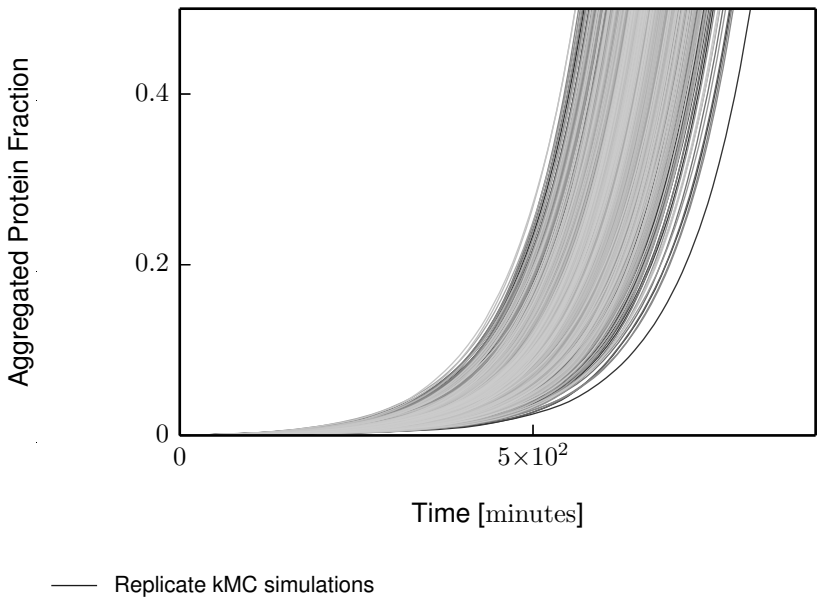


Figure 4.10 Growth curves produced by kMC simulations of the simplified, general autocatalytic polymerisation model described by reactions 4.9 and 4.10, where the parameter values were $\mathcal{N}_0 = 1$, $\mathcal{M}_0 = 2$, $\mu = 1\ \mathrm{s}^{-1}$, $\nu = 3{\times}10^{-8}\ \mathrm{s}^{-1}$. These $10,000$ replicate growth curves are only shown up to $\phi_M = 0.5$, since the model is only valid at "early times".

In the absence of monomer depletion, both $\mathcal{N}$ and $\mathcal{M}$ are free to grow indefinitely from their initial values $\mathcal{N}_0$ and $\mathcal{M}_0$. Growth curves from kinetic Monte Carlo simulation of this simplified, coupled growth model are shown in figure 4.10. In these simulations the monomer population is not depleted and so aggregation is only shown up to $\phi_M = 0.5$. These growth curves strongly resemble the early time portion of the growth curves for the full simulation presented in §3 for both the fragmentation and heterogeneous nucleation model, suggesting that this simple model is a realistic approximation for the early time behaviour of those models.

An analytic solution to a pair of simple coupled growth processes

An analytic expression describing the distribution of lag times for the growth curves in figure 4.10 was obtained by my collaborator Juraj Szavitz-Nossan. As the full derivation is not particularly illuminating to the results, I will only present a brief description here [(9)].

[(9)] A more detailed derivation can be found in §A3.

Using only reactions 4.9 and 4.10 [10], a general master equation for coupled growth can be written thus:

$$\frac{dP_{\mathcal{N},\mathcal{M}}}{dt} = \mathcal{M}\nu P_{\mathcal{N}-1,\mathcal{M}} + \mathcal{N}\mu P_{\mathcal{N},\mathcal{M}-1} - (\mathcal{M}\nu - \mathcal{N}\mu)P_{\mathcal{N},\mathcal{M}} \tag{4.11}$$

This can be solved using Van Kampen's system size expansion technique with the small parameter $1/v = 1/N_A V$ [59, 105], to obtain the probability distribution $P_{\mathcal{N},\mathcal{M}}(t)$ of finding a specific number of filaments with a specific total monomer content at time t. From this, the waiting-time distribution to reach M_ϕ can be obtained in terms of the more familiar number concentration $N(t) = \mathcal{N}(t)/v$ and fibrillar protein concentration $M(t) = \mathcal{M}(t)/v$:

$$L(t) = \frac{(2F(t)M'(t) + (M_\phi - M(t))F'(t))\sqrt{v}}{2\sqrt{2\nu\pi F^3(t)}} e^{-\frac{(M_\phi - M(t))^2 v}{2\nu F(t)}} \tag{4.12}$$

where $F(t)$ (given in §A3) is a function describing the coupled exponential growth process, and F' is its time derivative.

In the limit of large volume (i.e. where the Van Kampen approximation becomes exact), equation 4.12 can be approximated by a Gaussian distribution, giving an analytic form for the mean lag-time very similar in form to equation 1.20:

$$\mu_\tau = \frac{1}{\sqrt{\mu\nu}} \ln\left(\frac{M_\phi + \sqrt{{M_\phi}^2 + N^2(0)\frac{\mu}{\nu} - M^2(0)}}{N(0)\sqrt{\frac{\mu}{\nu}} + M(0)}\right) \tag{4.13}$$

This expression depends on the initial conditions $M(0)$ and $N(0)$, as well as the rate parameters μ and ν, all of which are independent of the system size. On the other hand, the variance of this Gaussian approximation scales inversely with the system size according to

$$\sigma_\tau^2 = \frac{\nu F(\mu_\tau)}{\mu^2 N^2(\mu_\tau) v} \tag{4.14}$$

Comparison of the simplified and full simulations, and analytic predication

Figure 4.11 shows the lag-time distributions obtained from kinetic Monte Carlo simulations of the full and simplified autocatalytic polymerisation models along

[10] I will ignore primary nucleation and depolymerisation in this simplified model.

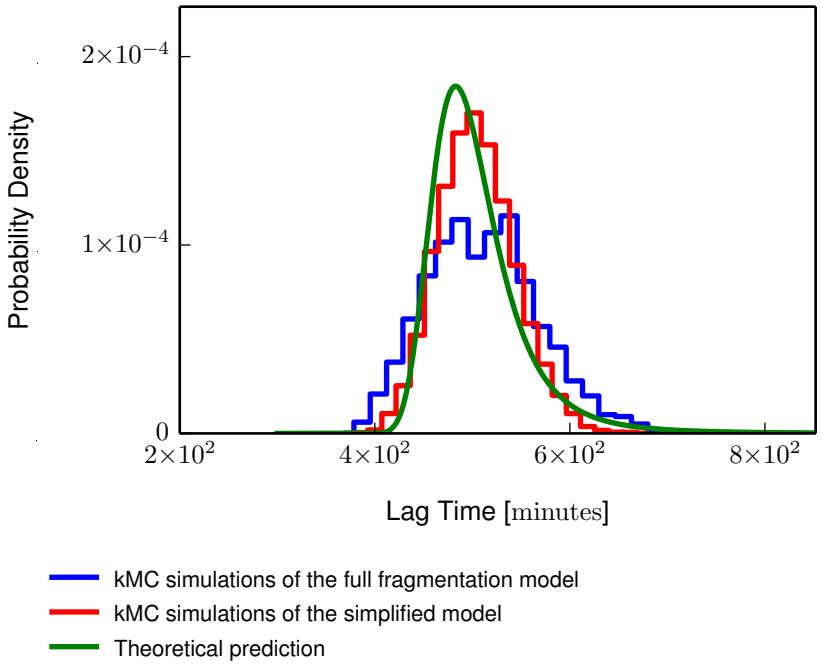


Figure 4.11 The theoretical probability distribution of lag times for autocatalytic polymerisation. The **histograms** contain lag times from replicate kMC simulations of the fragmentation model (previously shown in figure 4.9b), and simplified model (extracted from the replicate growth curves in figure 4.10); and the **curved line** is the theoretical prediction given by equation 4.12. The parameter values for the fragmentation model were $m_{\mathrm{tot}} = 10\ \mu\mathrm{M}$, $n_{\mathrm{c}} = 2$, $N(0) = 2{\times}10^{-12}\ \mathrm{M}$, $k_{\mathrm{n}} = 0$, $k_{+} = 5{\times}10^{4}\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$,$k_{-}$, $k_{\mathrm{f}} = 3{\times}10^{-8}\ \mathrm{s}^{-1}$, and $V = 83{\times}10^{-14}\ \mathrm{l}$, from which $\mathcal{N}_0$, $\mathcal{M}_0$, μ, and ν were derived.

with the prediction from equation 4.12. The level of agreement between the distributions produced by both kinetic Monte Carlo simulations is encouraging, and implies that the simplified model is able to capture a reasonable amount of the early time behaviour of the full system.

The coarse-grained simulation only considers a simplified form of the fragmentation mechanism: fragmentation should be proportional to the number of monomer-monomer bonds $(M(t) - N(t))$ in all filaments and not the number of monomers $(M(t))$ as I have assumed in the simple model. This will have resulted in an overestimate of the fragmentation probability. Additionally, monomer depletion was explicitly ignored in the coarse-grained simulations. These factors are probably responsible for the broader distribution produced by

the full simulation, as they both increase the number of steps necessary to reach the threshold; although their influence does not appear to be particularly large.

Equation 4.12 is also in reasonable agreement with the distribution produced by the simplified simulation but predicts a slightly lower mean lag-time. This indicates that the simulation either involves more steps than the model accounts for, or that each of these steps has a fractionally longer mean waiting-time than expected. The former could be caused by a discrepancy in the monomer count needed to reach ϕ_M, between the discrete simulation and continuum analytic solution. The latter, more likely explanation, is that monomer depletion is causing the rate of elongation to vary with time.

While this distribution successfully explains the lag-time distribution of the full kinetic Monte Carlo simulations, it does not provide an adequate explanation for the experimental lag-time distribution. From equation 4.14 it can be seen that the width of the lag-time distribution scales inversely with the system volume: predicting variance of the order of milliseconds for the volume of 100 μl (i.e. the volume used in the experimental results shown). This is clearly many order of magnitude smaller than the variability apparent in the experimental data and hence, autocatalysis alone cannot fully account for the experimental lag-time distribution. For this reason I will now discuss the effect that primary nucleation has on these predictions.

4.3.2 The effect of primary nucleation on the waiting-time distributions of an autocatalytic polymerisation model

When the rate of primary nucleation is much lower than that of autocatalysis, the time between consecutive nucleation events will be long, and only the first nucleus produced will be important to the kinetics: once this first nucleus is formed, the kinetics will be dominated by autocatalytic polymerisation such that all other nucleation events can be neglected. From this it can be seen that the total waiting-time can be accounted for simply by adding an appropriately distributed random time (representing the time at which the nucleus forms) to the waiting time for autocatalytic polymerisation, where the initial conditions for the later process are given by the primary nucleation event.

As in the simulations, I will assume that the formation of a primary nucleus has an exponentially distributed waiting-time

$$t_0 = \ln\left(\frac{1}{r}\right)^{N(0)/\eta} \tag{4.15}$$

where the primary nucleation rate is given by $\eta/N(0)$ (defined in §3.3.3), $N(0)$ is the concentration of nuclei produced and r is a random number drawn from the uniform distribution in the interval $(0, 1]$. For the formation of a single nucleus, the concentration of nuclei $N(0) = 1/(N_A V)$ is dependent on system volume, making it more likely that a nucleus forms in a larger volume where the number of monomers is correspondingly higher. This also defines the "initial conditions" for autocatalysis, $M(0) = n_c N(0)$ at t_0.

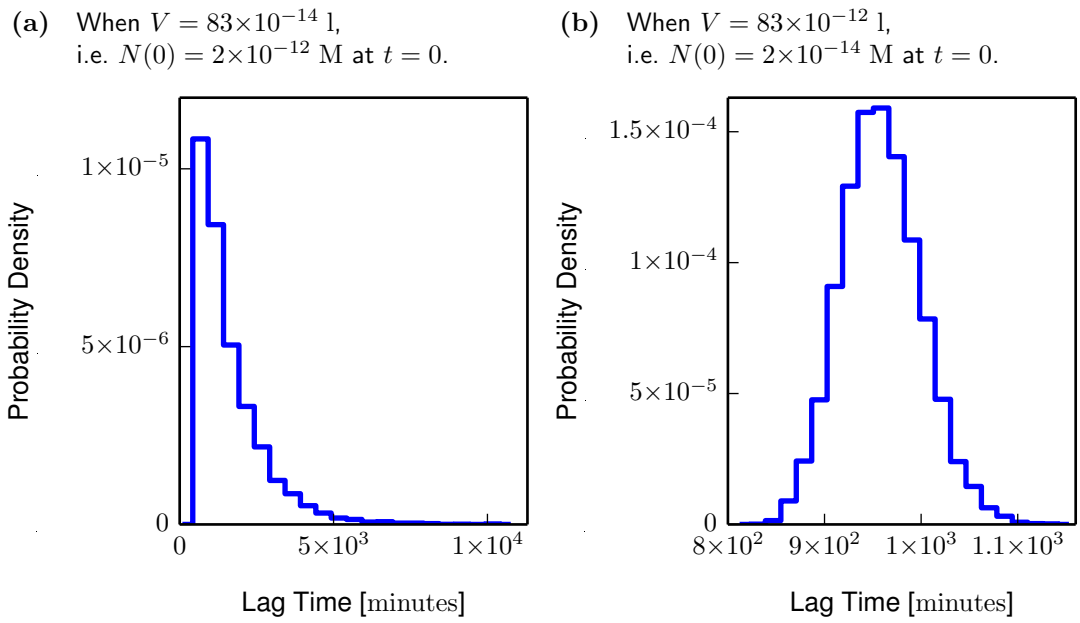


Figure 4.12 The distribution of lag times from kMC simulation growth curves of the simplified, general autocatalytic polymerisation model described by reactions 4.9 and 4.10, for two system volumes, V, where the first primary nucleation event is included as a start time t_0, randomly sampled using equation 4.15 for the seed concentrations, $N(0)$. The common parameter values were $m_{tot} = 10$ μM, $n_c = 2$, $k_+ = 5\times10^4$ M^{-1} s^{-1}, $k_- = 0$, and $k_f = 3\times10^{-8}$ s^{-1}. These distributions each contain lag times for $10,000$ replicate growth curves, where the lag time is defined as the point in the growth curve when 10% of the total protein concentration had been incorporated into filaments.

This method was used to produce the lag-time distributions shown in figure 4.12; combining the nucleation waiting time t_0 defined above, with the simulation results shown previously in figure 4.11.

From figure 4.12a it can be seen that, for the same volume and value of $N(0)$ as used in figure 4.11, the first primary nucleation event dominates the lag-time distribution. This agrees very well with the lag-time distribution produced by the full kinetic Monte Carlo simulation in which primary nucleation was active, indicating that primary nucleation was indeed responsible for the shape of that distribution (figure 4.2a).

Conversely, for a larger volume (shown in figure 4.12b) primary nucleation has a higher probability, and thus t_0 is small compared to the waiting time for autocatalysis. This case qualitatively recovers the peaked distribution of lag times produced by the full simulation of seeded autocatalytic polymerisation (figure 4.11): the higher mean lag-time and reduced variance in figure 4.12b is due to the larger volume, corresponding to a lower concentration of nuclei (as discussed in §3.3.3) and larger number of monomers than was present in the full simulation. Hence, in this case the lag time is dominated by autocatalysis, with primary nucleation only contributing to the value of the mean.

Since the lag-time distribution in figure 4.12b is dominated by autocatalytic polymerisation, its variance is likely to be described by equation 4.14, with only a minor contribution from primary nucleation. Thus, for a volume comparable with that used in the experimental results shown (i.e. 100 μl), the variability will still be vanishingly small, as will the contribution from the first primary nucleation event. Hence, this model shows that the experimentally observed lag-time variability cannot be accounted for by nucleated, autocatalytic polymerisation either.

As an aside, the above analysis only considered the possibility that a single nucleus forms. For a larger volume, however, it can be seen from equation 4.15 that the waiting time for the first nucleus will become vanishingly small, and thus it is possible that more than one nucleus will form within a short space of time (very early in the lag phase). If this is approximated by the "simultaneous" formation of a fixed concentration of nuclei (i.e. that $N(0)$ can be considered volume independent), then the same analysis can be repeated but without the

volume dependence in equation 4.15. Doing this, reveals that under these conditions primary nucleation will always dominate the lag time, producing a distribution similar to that shown in figure 4.12a. Hence, while this does produce a distribution with variability that does not decrease with system volume, the shape of the distribution does not match that of the experimental lag times shown in figure 4.1, again indicating that primary nucleation cannot be the prominent contribution to the lag phase.

4.4 Discussion

An understanding of the mechanisms responsible for the early stages of amyloid fibril self-assembly is of particular importance as control of these mechanisms presents the best opportunity to initiate or prevent this type of aggregation. Unfortunately, the early stages of self-assembly are also the hardest to probe experimentally and for this reason mechanisms involved at early times are often inferred from the behaviour of the kinetics at later times.

Variability in the lag time of replicate experimental growth curves has been thought to indicate fluctuations in the primary nucleation process [45, 48, 111] and thus provide insight into these early-time mechanisms. The analytic results of the minimal models that I presented in this chapter, predict that the fluctuations in the processes of primary nucleation and autocatalytic polymerisation scale inversely with system volume and become negligible at a volume comparable with typical experiments; hence, neither primary nucleation nor autocatalytic polymerisation, in the forms that I have investigated, can account for the variability in lag times observed experimentally.

Moreover, in almost all of the situations where primary nucleation constituted a slow step to self-assembly, the lag-time distributions were completely dominated by the waiting time for this one rare event. This resulted in a shifted exponential lag-time distribution — where the shift was caused by other, subsequent processes such as polymerisation or autocatalysis — with a very different appearance to the peaked distributions observed experimentally [45]. Thus, further indicating

that a single rare event is unlikely to be responsible for the lag-time distributions seen experimentally [(11)].

That said, I have only considered the formation of a primary nucleus by polymerisation or as a single-step process, and neglected depolymerisation in all but one case. Depolymerisation is often neglected, when seeking an analytic solutions, by assuming that the abundance of free monomer favours polymerisation at early times [6, 14, 15, 24, 55, 81, 82, 106]. However, given that this process is seen to actually increase the variability with increasing volume, it could be crucial at early times [12]. If short filaments are less stable and have a much higher depolymerisation rate than longer filaments, then the effects of depolymerisation are expected to be more apparent when filaments are short and few in number. Thus, I expect depolymerisation to have an important effect if taken into account in primary nucleus formation, or in the fragmentation model [(12)].

Amyloid fibril self-assembly is notorious sensitive to the environmental conditions and it has been suggested by Hellstrand et al. [43] that multiple factors may contribute to the variability observed experimentally [43, 111]. Indeed, on their own, the experimentally determined lag-time distributions (broad and approximately Gaussian) may provide very little information about the mechanisms, as this shape would be expected (from central limit theorem) for any sufficiently long sequence of random events; i.e. the sum of many steps, each with randomly distributed waiting-times. However, these distributions can be expected to scale with the system volume and so a better understanding of the mechanisms responsible for the variability might be found by comparing lag-time distributions for different volumes. A method by which this could be achieved has already been demonstrated by Knowles et al. [56] for the mean lag-time.

Finally, the effect that the system volume has on the waiting time, poses an interesting question about the comparison between laboratory experiments and those conducted in a volume comparable with a biological cell. The cellular environment typically contains a far higher protein concentration than is used experimentally, and in a much smaller and more crowded environment. Under

(11) Primary nucleation's dominance during the lag phase will become important in the next chapter where I explore the scaling of the mean lag-time with total protein concentration.

(12) It has been shown elsewhere that short filaments can have a different probability of breaking [44, 47].

such conditions aggregation might be expected to occur more readily and a number of studies do appear to show this [65, 76, 116]. However, while those studies used higher concentrations and "crowding agents" to simulate the cellular environment, they were still conducted in volumes of the order of $\sim \mu l$: under these conditions, my findings indicate that the probability of primary nucleation is expected to be very high. Hence, the aggregation propensity of proteins studied *in vitro* might be significantly lower *in vivo*, and caution should be exercised when comparing effects that might depend on volume, at such radically different volumes.

Chapter 5

The effects of primary nucleation on the scaling behaviour of the lag time with total protein concentration

5.1 Introduction

Most experiments on amyloid fibrils under typical laboratory conditions involve protein concentrations of the order of μM and volumes in the range ml– μl. These conditions mean that the system will contain a very large number of protein molecules and thus can be considered to be in the deterministic limit. However, volumes much closer to the cellular scale (of the order 10^{-10} l) have also been investigated [56]. At these scales, Knowles et al. [56] reported an increase in lag time with decreasing volume and speculated that this was due to a decrease in the probability of primary nucleation.

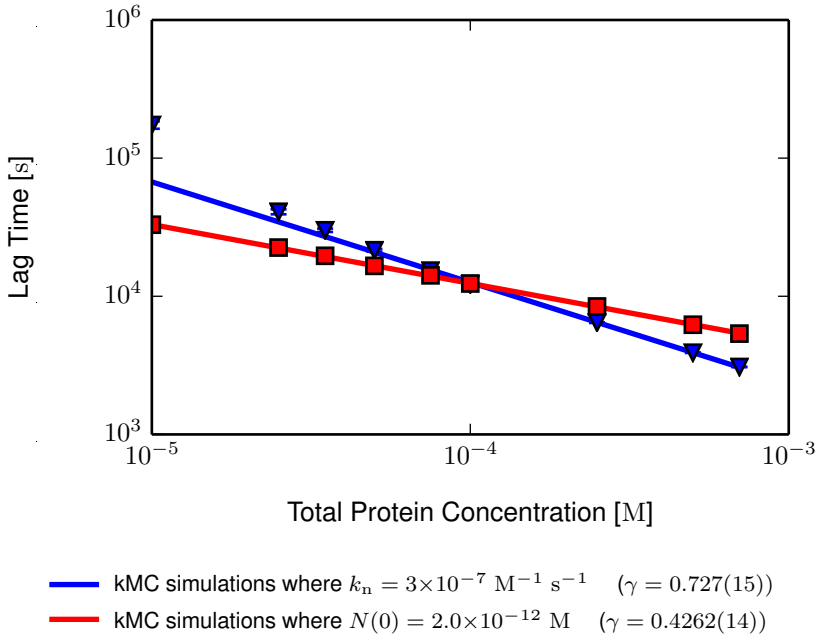


Figure 5.1 The scaling behaviour of τ_3 with m_{tot} in kMC simulations of the fragmentation model, when self-assembly either includes primary nucleation (with rate k_n) or is seeded (with $N(0)$ preformed filaments); where the common parameter values were $n_c = 2$, $k_+ = 5\times10^4$ M^{-1} s^{-1}, $k_- = 0$, $k_f = 3\times10^{-8}$ s^{-1}, and $V = 83\times10^{-14}$ l. Each kMC simulation point is the mean of 150 replicates with error bars indicating the standard deviation (method in §2.2); and the lines indicate the best fit of the power law $\tau_{\text{lag}} \propto m_{\text{tot}}^{-\gamma}$ to the data set (method in §2.3.2).

Using kinetic Monte Carlo simulations, I demonstrated in §3 and §4 that primary nucleation has a profound effect on the lag time and that this can be influenced by system size, either by affecting the waiting-time distribution for the nucleation event, or by changing the concentration of nuclei produced, or both. Simulation results for the lag-time scaling behaviour (discussed in §3) are shown again in figure 5.1: for the simulations where primary nucleation is active, it is apparent that the scaling behaviour deviates from the predicted power law at low protein concentration; i.e. at low protein concentrations, the lag times do not lie on a straight line on this log-log plot.

Typically a power law of the form

$$\tau_{\text{lag}} = A m_{\text{tot}}^{-\gamma} \tag{5.1}$$

is expected to describe the way that the lag-time scales with total protein concentration, where different models predict different values of the exponent γ, and A is an arbitrary scale factor independent of m_{tot}. Predictions for the exponent can be found by examining the protein concentration dependence of the characteristic time scale of the model at hand as follows.

The nucleation-dependent polymerisation model with characteristic time scale $\delta^{-1} = \left(2k_{+}k_{\mathrm{n}}m_{\mathrm{tot}}^{n_{\mathrm{c}}}/n_{\mathrm{c}}!\right)^{-1/2}$ predicts an exponent $\gamma = {}^{n_{\mathrm{c}}}/_{2} \geq 1$ for nuclei of size $n_{\mathrm{c}} \geq 2$, as was seen in §1.3.1. In a similar way, the lag-time expressions for the autocatalytic models presented in §1.3.3 all depend on the characteristic time scale κ^{-1}. This characteristic time scale is different for the two autocatalytic mechanisms presented: for the heterogeneous nucleation model $\gamma = {}^{(n_{\mathrm{s}}+1)}/_{2}$ for a secondary nucleus size $n_{\mathrm{s}} \geq 1$, while for the fragmentation model $\gamma = {}^{1}/_{2}$. These can both be expressed as $\gamma = {}^{(n_{\mathrm{s}}+1)}/_{2}$ where fragmentation has an effective secondary nucleus size of zero.

The fragmentation model is particularly important as it may explain the concentration scaling behaviour of a number of experimental systems, including the bovine insulin data presented earlier, which show scaling exponent values $\gamma < 1$. However, to obtain the analytic predictions for the scaling relations in the fragmentation and heterogeneous nucleation models, Cohen et al. [15] assumed that $\kappa \gg \delta$. In order to understand the results in figure 5.1 and determine how the scaling exponent changes with primary nucleation rate k_{n}, I will now explore the consequences of relaxing this assumption.

5.2 The influence of the primary nucleation rate on the lag-time scaling exponent in the fragmentation model

I extracted lag times (according to the method in §2.3.2) from kinetic Monte Carlo simulation growth curves of both fragmentation and heterogeneous nucleation models, produced (as described in §3) for a range of primary nucleation rates and total protein concentrations. The lag times from the fragmentation model are shown in figure 5.2 as a function of total protein concentration, for primary

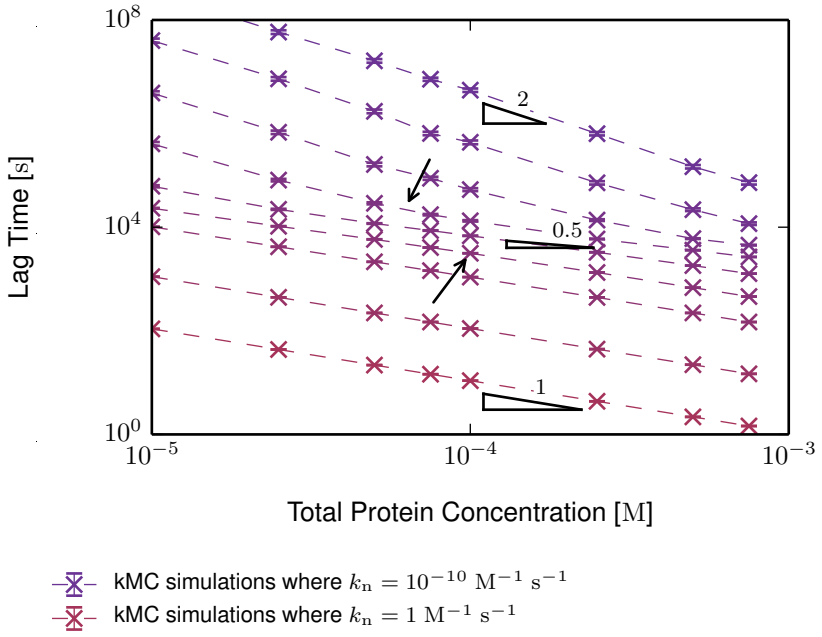


Figure 5.2 The scaling behaviour of τ_3 with m_{tot} in kMC simulations of the fragmentation model (described in §3), for different primary nucleation rates in the range $k_n = 10^{-15}$ M^{-1} s^{-1} to $k_n = 1$ M^{-1} s^{-1}, incremented by factors of 100; where the common parameter values were $N(0) = 0$, $n_c = 2$, $k_+ = 5{\times}10^4$ M^{-1} s^{-1}, $k_- = 0$, $k_f = 3{\times}10^{-8}$ s^{-1}, and $V = 83{\times}10^{-14}$ l. Each kMC simulation point is the mean of 150 replicates with error bars indicating the standard deviation (method in §2.2), and the lines are to help differentiate the data sets. The triangles illustrate specific scaling exponents and the arrows are to indicate the position of a "kink" where the gradient of the line changes.

nucleation rates in the range $10^{-10} \leq k_n \leq 1$ M^{-1} s^{-1} [(1)], with $n_c = 2$; the data for the heterogeneous nucleation model will be discussed later in §5.3.

From figure 5.2 it is evident that the primary nucleation rate k_n indeed affects the scaling behaviour and that the lag times for which $\gamma < 1$ appear sandwiched between two other regimes where $\gamma \geq 1$, as indicated by the triangles in figure 5.2. This contour plot also shows the "kink" hinted at in figure 5.1, where the scaling behaviour changes with changing protein concentration for a fixed value of k_n, and that the position of this kink is different for different primary nucleation rates.

(1) The units for the primary nucleation rate should be $M^{-(n_c-1)}$ s^{-1}, unless n_c is defined. In this chapter I look at cases where $n_c = 2, 4,$ and 8.

Another, less prominent kink can also be seen at higher protein concentrations for some primary nucleation rates, but was not apparent in figure 5.1.

Since values of γ must be obtained by fitting the power law equation 5.1 to data over a range of protein concentrations, the changes in gradient of the data for constant k_n shown in figure 5.2, will influence the value obtained for a given value of k_n. The effect which changing the range of protein concentrations has on the value of γ obtained, will be discussed later in §5.2.4, but for now, I will only use lag times produced in the range of protein concentrations shown in figure 5.2.

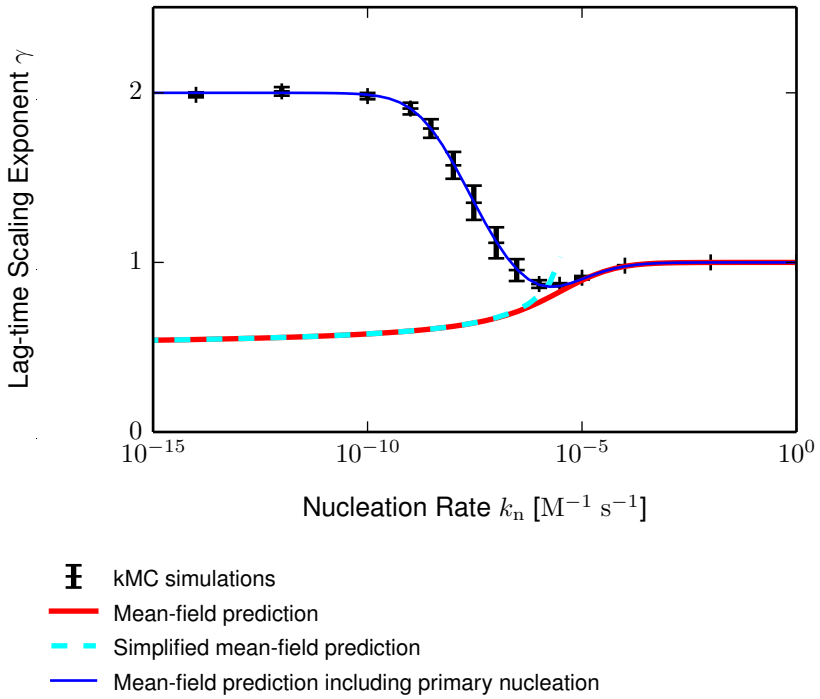


Figure 5.3 Theoretical predictions of the lag-time scaling exponent for different primary nucleation rates. The values of γ were determined by fitting lag times for total protein concentrations in the range 10–750 μM, to the expression $\tau_{lag} \propto m_{tot}^{-\gamma}$, where the lag times were obtained from: the mean τ_3 of 150 replicate growth curves of kMC simulations of the fragmentation model, with error bars indicating the confidence in the fitted value (i.e. the lag-time variance was not considered in the fit); and evaluation of the mean-field equation 5.2, the simplified mean-field equation 5.3, and the modified mean-field equation 5.4. The common parameter values were $N(0) = 0$, $n_c = 2$, $k_+ = 5{\times}10^4$ M^{-1} s^{-1}, $k_- = 0$, $k_f = 3{\times}10^{-8}$ s^{-1}, and $V = 83{\times}10^{-14}$ l.

Using the method presented in §2.3.2, I obtained values of the scaling exponent γ from kinetic Monte Carlo simulations over a range of primary nucleation rates

(shown as black points in figure 5.3). These encompass a larger range of values of k_n than were shown in figure 5.2, so as to completely cover the transition behaviour that I discuss below. It should be noted that the scaling exponent of these simulation data is always higher than the mean-field prediction for the fragmentation model (i.e. $\gamma > 1/2$); indeed it reaches a surprisingly large value ($\gamma = 2$) when the rate of primary nucleation is low and autocatalysis is expected to dominate. In order to understand this, the predictions for γ made by the various mean-field models outlined in §1.3.4, need to be examined in more detail.

5.2.1 Analytic predictions for the lag-time scaling exponent

Equation 1.20, which I presented previously in §1.3.4, describes the lag time to reach a threshold fraction of aggregate M_ϕ for a general model of autocatalytic polymerisation. This is given again for reference:

$$\tau_{\phi_M} = \frac{1}{\kappa} \ln \left(\frac{D - \Phi + \sqrt{(\Phi - D)^2 + 4C_+C_-}}{2C_+} \right) \tag{5.2}$$

where $D = \delta^2/\kappa^2$ and $\Phi = \ln(1 - \phi_M)$.

A simpler form of this expression was also presented in equation 1.21, which is also given for reference:

$$\tau_{\phi_M} = \frac{\log\left(\frac{\phi_M}{C_+}\right)}{\kappa} \tag{5.3}$$

In order to obtain γ from these lag-time expressions for an arbitrary value of k_n, it is necessary to use the same process as for the simulation data above, of fitting to multiple lag times generated over a rage of protein concentrations. This is because m_{tot} appears in multiple places throughout each of these expressions, in conjunction with both autocatalysis and primary nucleation (through κ and δ, respectively). By altering the primary nucleation rate, the relative contribution of these terms will change, affecting the concentration dependence of the entire expression. Only in limiting cases is it possible to obtain the values of γ from these expressions analytically, as I will show later in §5.2.1.

The dashed line in figure 5.3 was produced from the lag times generated by the simplified expression (equation 5.3). This line does come close to the predicted value of $\gamma = {}^1/_2$ when the primary nucleation rate is very low, but the equation fails when the rate becomes too high. This is unsurprising, as the expression was obtained from the more complicated equation 5.2 by assuming $\delta \ll \kappa$, and it is this assumption which fails when $k_\mathrm{n} \gtrsim 10^{-5}\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$. Hence, the lag times produced by equation 5.2 (shown as a solid red line in figure 5.3) agree with the simplified expression when the rate of primary nucleation is low, and are also able to accurately describe the simulation data when the rate of primary nucleation is high. It does not, however, reproduce the scaling behaviour apparent in the kinetic Monte Carlo simulation data when the rate of primary nucleation is low; the reason for this will be explained below.

Two regimes in which primary nucleation dominates the lag time

At a high rate of primary nucleation the kinetics are expected to be described by the nucleation-dependent polymerisation model where the scaling exponent is predicted to be $\gamma = {}^{n_\mathrm{c}}/_2$ for $n_\mathrm{c} \geq 2$ (see §1.3.1). This is indeed the case in figure 5.3, and is the reason for the higher scaling exponent value, seen in the data in the lower half of figure 5.2. The value of k_n where γ starts to decrease from $\gamma = 1$, is where the propensities of autocatalysis and primary nucleation are comparable. The conditions for this transition will be described in the next section.

The strong deviation of the γ values produced by the kinetic Monte Carlo simulations, from those predicted by equation 5.2 for the fragmentation model — as evident on the left hand side of figure 5.3 — can be understood by considering the competition between primary nucleation and autocatalytic polymerisation when the rate of primary nucleation is low. If $\delta \ll \kappa$, then autocatalytic polymerisation is expected to out-compete primary nucleation and dominate the kinetics. However, autocatalytic polymerisation cannot operate until the first primary nucleus has formed (as described previously in §4). Consequently the formation of the first primary nucleus is critical to all subsequent autocatalytic polymerisation. As I have already stated, primary nucleation is slow in this regime, so its contribution to the lag time can be treated independently from that of autocatalytic polymerisation. To do this I use a similar argument to the

one in §4.3; here, I first consider the time taken to form the first nucleus, and then separately the lag time for seeded autocatalytic polymerisation with a seed concentration $N(0)$ as determined by the first nucleation event.

As in §4, every primary nucleation event can be treated as a Poisson process with rate $\eta/N(0)$, where $\eta = k_{\mathrm{n}} m_{\mathrm{tot}}^{n_{\mathrm{c}}}/n_{\mathrm{c}}!$ is the propensity given in reaction 1.6 and $N(0) = M(0)/n_{\mathrm{c}} = 1/N_{\mathrm{A}}V$ is the number concentration of nuclei produced per volume V. From this, the mean waiting-time $\tau_{\mathrm{Poisson}} = N(0)/\eta$ can be obtained; i.e. the mean-field behaviour of primary nucleation [(2)]. Provided $n_{\mathrm{c}}N(0) < M_\phi$ then the waiting time for the first nucleation event can simply be added to that of autocatalytic polymerisation (equation 1.20), to give a complete mean-field prediction for the lag time:

$$\tau_{\phi_M} = \frac{N(0)}{\eta} + \frac{1}{\kappa} \log\left(1 + \Phi D^{-1} + \sqrt{\Phi^2 D^{-2} + 2\Phi D^{-1}}\right) \tag{5.4}$$

where $D = \delta^2/\kappa^2$ and $\Phi = -\log(1 - \phi_M)$ as before, and the logarithms are base e.

Since $\eta = \delta^2/2k_+$, equation 5.4 predicts that when $\delta \ll \kappa$, necessarily $\tau_{\phi_M} \ll \tau_{\mathrm{Poisson}}$, and the concentration scaling will be dominated by τ_{Poisson}, giving $\gamma = n_{\mathrm{c}}$. Once the rate of primary nucleation becomes high enough, τ_{Poisson} becomes negligible compared to τ_{ϕ_M}, even if only a single primary nucleus is formed, and the scaling changes to that of autocatalytic polymerisation. The solid blue line in figure 5.3 shows the prediction of equation 5.4, which describes the kinetic Monte Carlo simulation data extremely well over the entire range of primary nucleation rates.

Thus, the picture which emerges from the simulations is one in which primary nucleation dominates the lag-time scaling of the fragmentation model when the rate at which it occurs is either very low or very high. Lag times produced by equation 5.4 for different primary nucleation rates are shown in figure 5.4, along with the simulation data from figure 5.2. This plot makes a clear prediction for experiments: if lag times are obtained over a wide enough range of protein concentrations, these will not be described by a single power-law, but rather exhibit a number of transitions between power-laws, each with a different value of the scaling exponent. The transitions between these different scaling exponents

[(2)] A missing mean-field description of primary nucleation was also the cause of the discrepancy between the average of the simulation growth curves and the mean-field prediction, in figures 3.4 and 3.5 (§3.3.2).

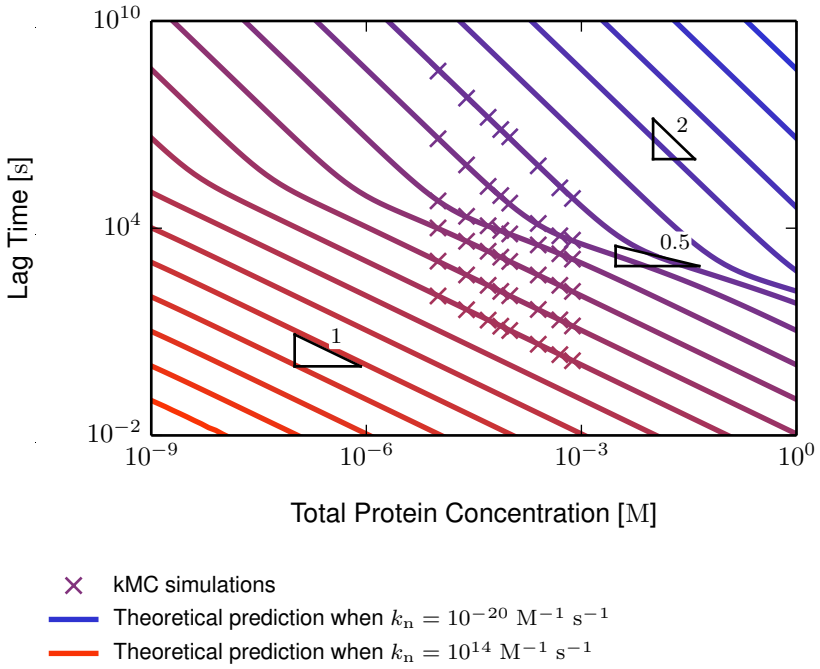


Figure 5.4 Theoretical prediction of equation 5.4, for the scaling behaviour of τ_{ϕ_M} with m_{tot} in the fragmentation model, with different primary nucleation rates in the range $k_n = 10^{-20}$ M^{-1} s^{-1} to $k_n = 10^{14}$ M^{-1} s^{-1}, incremented by factors of 100; where the common parameter values were $N(0) = 0$, $n_c = 2$, $k_+ = 5{\times}10^4$ M^{-1} s^{-1}, $k_- = 0$, $k_f = 3{\times}10^{-8}$ s^{-1}, and $V = 83{\times}10^{-14}$ l. Also shown are the kMC simulation data from figure 5.2. The triangles illustrate specific scaling exponents.

will manifest as kinks when the lag times are plotted as a function of protein concentration, as in figure 5.4, and are caused by a change in the mechanism dominating the lag phase (either of the two primary nucleation regimes, or that of autocatalytic polymerisation).

The conditions for the transitions between the regime dominated by autocatalysis and either of those dominated by primary nucleation can be obtained analytically by inspecting limiting cases of equation 5.4, as I will now show.

Conditions for the transitions between autocatalysis dominated and primary-nucleation dominated lag-time scaling regimes

When k_n is large, primary nucleation dominates the kinetics and the first term in equation 5.4 can be neglected. This allows a series expansion in $\mathcal{O}(\Phi D^{-1})$ of the logarithm, where Φ will always be vanishingly small when ϕ_M is small, giving

$$\lim_{D \gg \Phi} \tau_{\phi_M} = \frac{\sqrt{2\Phi}}{\delta} \tag{5.5}$$

which recovers the observed scaling $\gamma = {}^{n_\mathrm{c}}/_{2}$. The condition $D \gg \Phi$ can thus be used to obtain the value of $k_\mathrm{n} > \Phi k_\mathrm{f} m_\mathrm{tot}^{1-n_\mathrm{c}} n_\mathrm{c}!$ necessary for primary nucleation to dominate the lag phase. For values of k_n smaller than this, either autocatalytic polymerisation or slow primary nucleation will dominate the lag time, as I now show.

When k_n is small, the first term in equation 5.4 becomes large and the dominant term in the logarithm is ΦD^{-1}; thus I can write

$$\lim_{D \ll \Phi} \tau_{\phi_M} = \frac{1}{\delta^2} \left[2k_+ N(0) - D\kappa \log\left(\Phi^{-1} D\right) \right] \tag{5.6}$$

The two terms ${}^{2k_+ N(0)}/_{\delta^2}$ and ${}^{D\kappa \log\left(\Phi^{-1} D\right)}/_{\delta^2}$ in this expression scale differently with k_n, with important consequence for the lag-time scaling.

As $k_\mathrm{n} \to 0$, D will vanish faster than the rate at which $\log\left(\Phi^{-1} D\right) \to -\infty$ [3]. Because of this, the second term in equation 5.6 vanishes, recovering $\gamma = n_\mathrm{c}$; i.e. primary nucleation-dominated lag-time scaling.

When the primary nucleation rate is not quite so low, but still in the limit described by equation 5.6, the first term in the equation vanishes, recovering $\gamma = {}^1/_2$ as predicted by the autocatalysis model for fragmentation. This is conditional on the primary nucleation rate being

$$k_\mathrm{n}^* < k_\mathrm{n} < \Phi \frac{\kappa^2 n_\mathrm{c}!}{2k_+ m_\mathrm{tot}^{n_\mathrm{c}}}$$

[3] This can be shown from the definition $\lim_{x \to 0} x \log(x) = 0$.

where k_{n}^* is the value of the nucleation rate at the transition and is given by the condition that both terms in equation 5.6 make a comparable contribution to the lag time:

$$\frac{N(0)\kappa n_{\mathrm{c}}!}{k_{\mathrm{n}}^* m_{\mathrm{tot}}^{n_{\mathrm{c}}}} \approx \log\left(\frac{2k_+ k_{\mathrm{n}}^* m_{\mathrm{tot}}^{n_{\mathrm{c}}}}{\Phi\kappa^2 n_{\mathrm{c}}!}\right) \tag{5.7}$$

The value of k_{n}^* needs to be extracted from equation 5.7 numerically, but even without doing so it can be seen from the above condition on k_{n} that if $k_{\mathrm{n}}^* \geq \Phi\dfrac{\kappa^2 n_{\mathrm{c}}!}{2k_+ m_{\mathrm{tot}}^{n_{\mathrm{c}}}}$ primary nucleation will always dominate the lag time and there will be no regime in which $\gamma = 1/2$ is found. This is indeed seen in figure 5.4, where the two transition kinks have merged when the rate of primary nucleation is high.

In summary, the transitions between the different scaling behaviours are defined by:

$$k_{\mathrm{n}} \geq \Phi\frac{\kappa^2 n_{\mathrm{c}}!}{2k_+ m_{\mathrm{tot}}^{n_{\mathrm{c}}}} \qquad \text{when rapid primary nucleation dominates the entire growth curve;} \tag{5.8a}$$

$$k_{\mathrm{n}}^* < k_{\mathrm{n}} < \Phi\frac{\kappa^2 n_{\mathrm{c}}!}{2k_+ m_{\mathrm{tot}}^{n_{\mathrm{c}}}} \qquad \text{when autocatalysis dominates the entire growth curve (including the lag-time); and} \tag{5.8b}$$

$$k_{\mathrm{n}} \leq k_{\mathrm{n}}^* \qquad \text{when slow primary nucleation dominates the lag time, but autocatalysis dominates the subsequent growth;} \tag{5.8c}$$

where k_{n}^* is given by the solution to equation 5.7.

5.2.2 How the system volume alters the scaling behaviour observed

The results that I have presented so far, have been obtained for a volume of $V = 0.83$ pl. I will now show how changing this volume affects those results, and in particular how it alters the region of parameter space for which autocatalysis dominates the lag time.

Since equation 5.2 was obtained from the mean-field expression 1.14, the only volume dependence in equation 5.4 comes from the term for the first primary nucleation event, for which the waiting-time scales inversely with the system volume V in which the single nucleus forms. In larger volumes with the same protein concentration, the probability of nucleation will be higher, and the mean waiting-time to the first primary nucleation event, shorter. Consequently, changing V will only affect the region of parameter space in which self-assembly is dominated by the slow formation of the first primary nucleus.

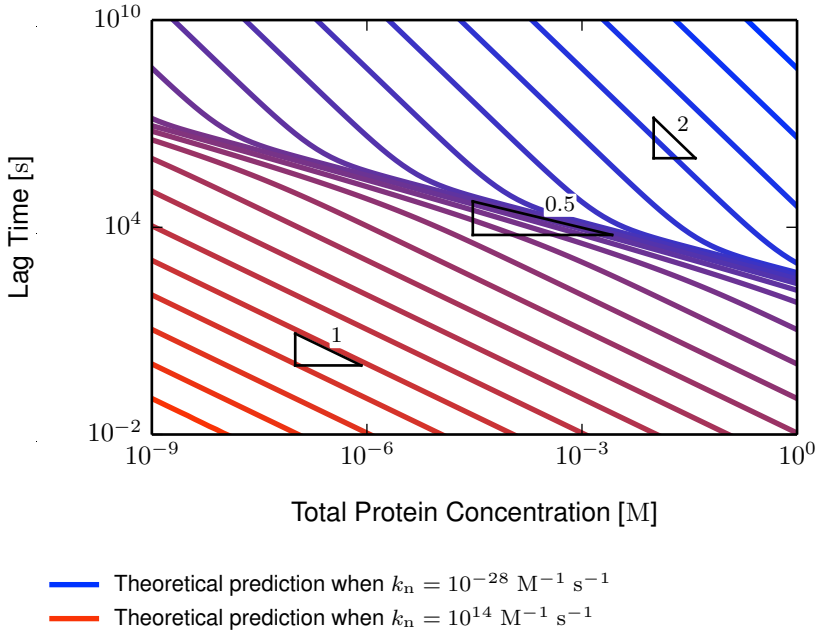


Figure 5.5 Theoretical prediction of equation 5.4, for the scaling behaviour of τ_{ϕ_M} with m_{tot} in the fragmentation model, with different primary nucleation rates in the range $k_n = 10^{-28}$ M^{-1} s^{-1} to $k_n = 10^{14}$ M^{-1} s^{-1}, incremented by factors of 100; where the common parameter values were $N(0) = 0$, $n_c = 2$, $k_+ = 5{\times}10^4$ M^{-1} s^{-1}, $k_- = 0$, $k_f = 3{\times}10^{-8}$ s^{-1}, and $V = 83{\times}10^{-6}$ l (larger than in figure 5.4). The triangles illustrate specific scaling exponents.

Lag times predicted by equation 5.4 for a volume 10^8 times larger than in figure 5.2, are shown in figure 5.5 [(4)]. In this, the three different scaling regimes are now more distinct, with the upper right portion dominated by slow nucleation, and the lower left by fast nucleation.

(4) Here, $V = 83$ µl is closer to a typical, *in vitro* experimental volume.

The larger volume has decreased the range of values of k_n over which slow nucleation dominants the lag phase, and thus extended the crossover region where $\gamma = 1/2$ is now evident. This can be understood from equation 5.7, which states that when a small concentration of filaments is produced by the first nucleation step, the value of k_n^* will also be small and thus increase the range of values of k_n for which condition 5.8b, holds. Thus, increasing the volume extends the range of total protein concentrations for which autocatalytic polymerisation ($\gamma = 1/2$) is dominant. This is interesting because it suggests that $\gamma = 1/2$ scaling may be more prominent in laboratory experiments, which are typically performed in large volumes, than in situations where the volume is comparable to that of a biological cell.

5.2.3 The influence of primary nucleus size

In both of the regions where primary nucleation dominates the lag time (i.e. high and low nucleation rates), the primary nucleus size n_c affects γ by altering the concentration dependence of the primary nucleation rate. Figure 5.6 shows that, by increasing n_c, the curves for γ as a function of k_n, shift towards higher k_n while (qualitatively) maintaining their shape. Additionally, the range of k_n for which γ is a minimum becomes narrower as n_c increases: with increased n_c, k_n^* becomes larger and the first primary nucleation event dominates the lag time for a broader range of k_n values. However, it would be extremely difficult to use this to determine n_c with any confidence, without examining the scaling behaviour of a wide range of protein concentrations.

5.2.4 How the choice of protein concentrations can influence which scaling behaviour is found

From the contour plots in figures 5.4 and 5.5 it is easy to see that, at different protein concentrations, different values of k_n are required in order to obtain the same value of γ; i.e. to obtain $\gamma = 1/2$ at high protein concentrations, the primary nucleation rate needs to be lower than it would be for a lower protein concentration.

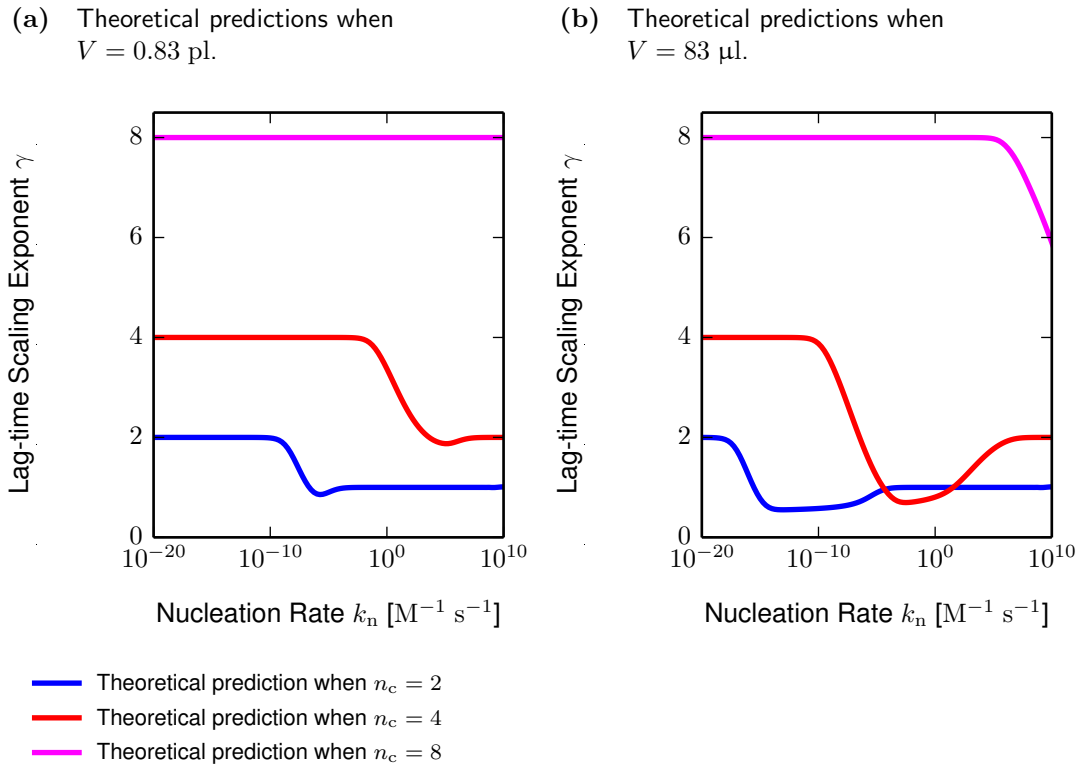


Figure 5.6 Theoretical prediction of equation 5.4, for the lag-time scaling exponent at different primary nucleation rates, for three sizes of primary nucleus, n_c, if the system volume is **(a)** the same as in figure 5.4, and **(b)** the same as in figure 5.5. The values of γ were determined by fitting lag times for total protein concentrations in the range 10–750 µM, to the expression $\tau_{lag} \propto m_{tot}^{-\gamma}$. The common parameter values were $N(0) = 0$, $k_+ = 5{\times}10^4$ M^{-1} s^{-1}, $k_- = 0$, and $k_f = 3{\times}10^{-8}$ s^{-1}.

The total protein concentration affects equation 5.4 by altering both η and δ, but also appears in κ. This means that the relative contribution of autocatalysis — compared to that of primary nucleation — is affected. Hence, both the range and value of concentrations over which the lag time is measured will affect the scaling behaviour seen for a specific primary nucleation rate. This dependence of the scaling behaviour on the protein concentration range is currently not widely recognized in the literature. Since the lag-time scaling exponent is often used to identify the mechanisms responsible for amyloid fibril self-assembly, this could lead to incorrect assumptions about the mechanisms involved.

To illustrate this effect, the values of γ obtained from the contours in figure 5.4 for two distinct ranges of total protein concentration are shown in figure 5.7. I

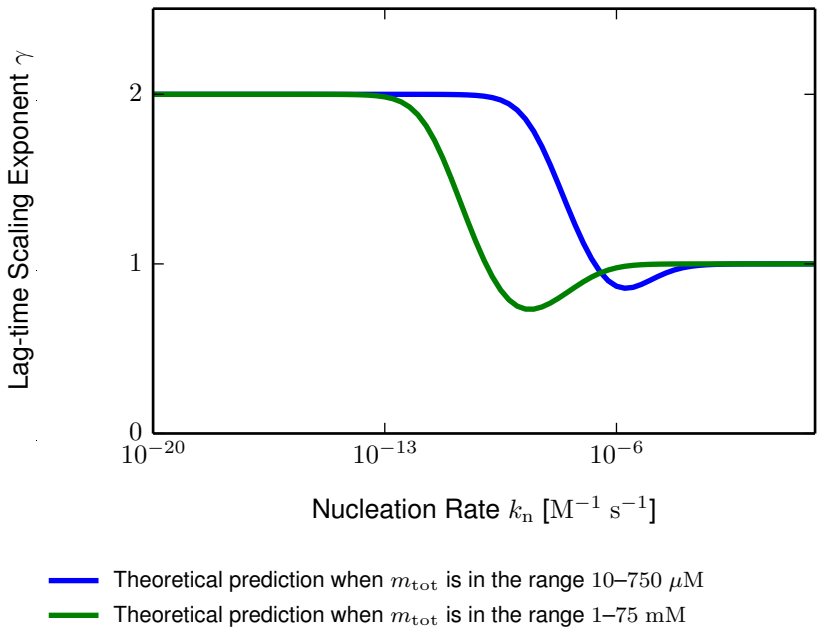


Figure 5.7 Theoretical prediction of equation 5.4, for the lag-time scaling exponent at different primary nucleation rates, when a different range of total protein concentrations is used to determine γ: the values of γ were determined by fitting lag times for the two total protein concentrations ranges, to the expression $\tau_{\mathrm{lag}} \propto m_{\mathrm{tot}}^{-\gamma}$. The common parameter values were $N(0) = 0$, $n_{\mathrm{c}} = 2$, $k_+ = 5{\times}10^4\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$, $k_- = 0$, $k_{\mathrm{f}} = 3{\times}10^{-8}\ \mathrm{s}^{-1}$, and $V = 83{\times}10^{-14}$ l.

elected to use the smaller volume to illustrate this because of the way that at lower total protein concentrations, the two kinks merge. Figure 5.7 demonstrates that by choosing a particular range of protein concentrations, the value of k_{n} for which the transitions between primary nucleation dominated and autocatalytic dominated lag times are observed, will shift, and the minimum value of γ may also change.

5.3 The influence of the primary nucleation rate on lag-time scaling exponent in the heterogeneous nucleation model

Thus far in this chapter, I have only discussed the modified lag-time expression (equation 5.4) in terms of the fragmentation model. However, this model is completely generalisable and can be easily extended to described the heterogeneous nucleation model as well; to do this I simply use the appropriate definition of κ (equation 1.18 from §1.3.3).

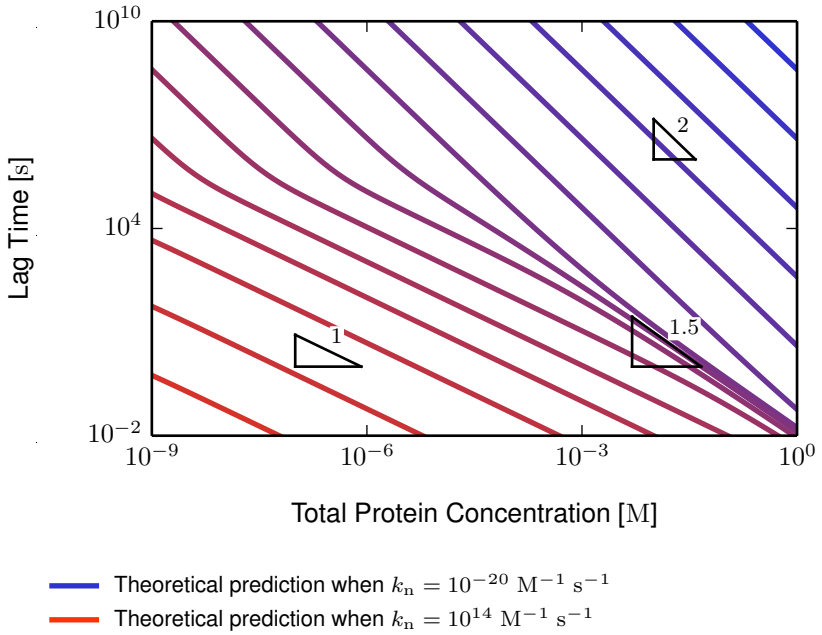


Figure 5.8 Theoretical prediction of equation 5.4, for the scaling behaviour of τ_3 with m_{tot} in the heterogeneous nucleation model, with different primary nucleation rates in the range $k_n = 10^{-28}$ M^{-1} s^{-1} to $k_n = 10^{14}$ M^{-1} s^{-1}, incremented by factors of 100; where the common parameter values were $N(0) = 0$, $n_c = 2$, $k_+ = 5{\times}10^4$ M^{-1} s^{-1}, $k_- = 3{\times}10^{-8}$ s^{-1}, $k_h = 24$ M^{-1} s^{-1}, $n_s = 2$, and $V = 83{\times}10^{-14}$ l. The triangles illustrate specific scaling exponents.

Figure 5.8 shows the lag times predicted by equation 5.4 if this deffition for κ is used; i.e. the heterogeneous nucleation with a secondary nucleus size $n_s = n_c = 2$. These can be directly compared with figure 5.4, since both plots are the same in all regards except the method of autocatalysis.

The scaling exponent predicted for the heterogeneous nucleation-dominated model is $\gamma = {}^{(n_s+1)}/_2$, and can indeed be seen in figure 5.8 at higher protein concentrations. For the value of n_s used, this γ value lies between those produced by the two primary nucleation dominated regimes [(5)] unlike in the fragmentation model where it was lower than the exponent of either primary nucleation-dominated regime [(6)].

Interestingly, in the fragmentation model, the scaling exponent of $\gamma = {}^1/_2$ is transient; at low protein concentrations the lag time is dominated by slow primary nucleation, whereas at high protein concentrations it is dominated by rapid primary nucleation, with the regime dominated by fragmentation sandwiched between these. This is not the case for the heterogeneous nucleation model when $n_s = n_c$ because, at high protein concentrations, the autocatalytic formation of filaments will occur more rapidly than primary nucleation. The effect can be seen better in figure 5.9, where a larger system volume has been used; once again, a larger system volume increases the range of total protein concentrations over which a γ value indicative of autocatalysis can be found.

The condition $n_s = n_c$, however, is a special case: when $n_c \geq n_s + 1$, a sufficiently high primary nucleation rate will always dominate the lag time, regardless of the protein concentration; whereas, when $2n_c \leq n_s + 1$, a sufficiently low rate of primary nucleation will always dominate the lag time, also regardless of the protein concentration. This can be understood analytically from conditions 5.8a, b, and c, by substituting the appropriate κ, and also graphically from the convergence of the lines of constant k_n in figures 5.8.

(5) The scaling exponent for lag times produced by the heterogeneous nucleation with $n_s = 2$, is $\gamma = {}^3/_2$, which lies between $\gamma = 2$ and $\gamma = 1$ for the slow and rapid primary nucleation regimes with $n_c = 2$, respectively.

(6) The scaling exponent for lag times produced by the fragmentation model is $\gamma = {}^1/_2$, which is less than both $\gamma = 2$ and $\gamma = 1$ for the slow and rapid primary nucleation regimes with $n_c = 2$, respectively.

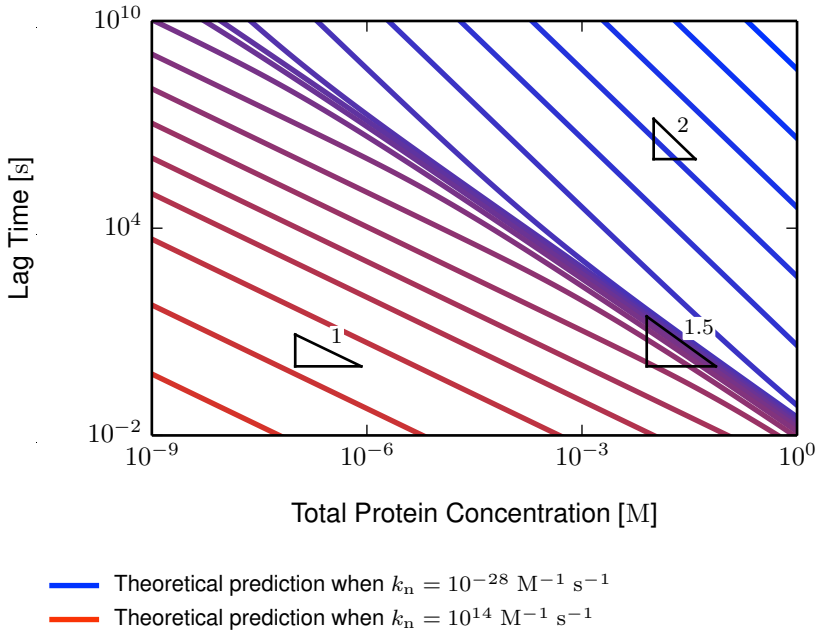


Figure 5.9 Theoretical prediction of equation 5.4, for the scaling behaviour of τ_3 with m_{tot} in the heterogeneous nucleation model, with different primary nucleation rates in the range $k_{\mathrm{n}} = 10^{-28}$ M^{-1} s^{-1} to $k_{\mathrm{n}} = 10^{14}$ M^{-1} s^{-1}, incremented by factors of 100; where the common parameter values were $N(0) = 0$, $n_{\mathrm{c}} = 2$, $k_+ = 5{\times}10^4$ M^{-1} s^{-1}, $k_- = 3{\times}10^{-8}$ s^{-1}, $k_{\mathrm{h}} = 24$ M^{-1} s^{-1}, $n_{\mathrm{s}} = 2$, and $V = 83{\times}10^{-6}$ l (larger than in figure 5.8). The triangles illustrate specific scaling exponents.

5.4 Discussion

The models for primary nucleation and autocatalysis predict different scaling behaviour of the lag time with total protein concentration, and this has been used as a means to identify which of the processes controls the kinetics [17]. However, the scaling behaviours predicted by some models may be indistinguishable when considered independently from other aspects of the kinetics [63]. Moreover, there may be scenarios in which the scaling behaviour is actually ambiguous as I will now discuss.

In this chapter I presented simulations and theoretical analysis which predict that, if measured over a wide range of protine concentrations, plots of lag time as a function of protein concentration should show "kinks" where there is a change in the mechanism dominating the lag phase. Such kinks have not previously

been discussed in the literature but lag-time data has been published which may indicate their presence [43, 87, 110, 111]. In other cases, the kinks may occur at protein concentrations that are not accessible experimentally, which means that the scaling behaviour will remain the same over all concentrations considered. However, if only a narrow range of protein concentrations is examined, these kinks could be missed, or dismissed as fluctuations in the data, which could explain why they have not been reported previously. Consequently, fitting a single power law to lag-time data taken from a too narrow a range of protein concentrations can result in an incorrect value of the scaling exponent by unwitting inclusion of part of one or both transitions.

Moreover, I demonstrated that under certain conditions the scaling behaviour predicted by autocatalysis cannot be seen at all, even though autocatalytic mechanisms are present in the model. At low primary nucleation rates or low protein concentrations, the waiting time to the formation of the first filament is significantly longer than the time scales of the rest of the kinetics, resulting in lag times dominated by primary nucleation even though the growth curves are otherwise dominated by autocatalysis (as demonstrated in figures 3.4 and 3.5). Consequently, the absence of lag-time scaling behaviour predicted by mechanisms such as autocatalysis over the range of protein concentrations examined does not necessarily mean that the kinetics are not otherwise dominated by these mechanisms. Therefore, I would advise examining as wide a range of protein concentrations as is experimentally feasible before attempting to extract a scaling exponent.

Curiously, the scaling exponent for bovine insulin [(7)] is lower than the minimum value predicted by the autocatalysis model (i.e. less than $\gamma = 1/2$ for the fragmentation model). Moreover, scaling exponents of this kind (i.e. $\gamma < 1/2$) have been consistently reported in the literature [18, 27, 53, 55, 74, 83, 94, 117], and even more may be evident if kinks are present as I speculated earlier [43, 87, 110, 111]. These lower values of the scaling exponent cannot be explained by the models that I have discussed here, and may represent another, as yet unknown, aggregation process with yet another scaling behaviour.

(7) The lag times for bovine insulin were given in figure 1.8.

As seen in the previous chapter, the system size again determines which process will dominate the lag phase and hence the conditions under which scaling behaviour indicative of autocatalysis can be observed. At very small volumes and low protein concentrations, only the primary nucleation dominated scaling behaviours will be observed. This is because the average waiting-time for the first primary nucleus is much longer than the lag-phase defined by autocatalysis alone when the rate of primary nucleation is low (see §4). This implies that even in the crowded cellular environment, the confinement of necessarily high concentrations of protein into small volumes may help prevent the initial stages of aggregation, thereby greatly increasing pathogenic latency of hereditary amyloidosis [56]. Such effects are likely to be missed by most laboratory experiments which utilise much larger volumes [11, 65, 76, 108, 116]. Hence, caution should be exercised when drawing comparison between these experiments and *in vivo* behaviour [100], since rare events will be much more influential in the latter but almost negligible in the former.

Chapter 6

The effects of arresting fragmentation during autocatalytic self-assembly

6.1 Introduction

The general autocatalytic polymerisation model presented in §1.3.2 predicts that the maximum aggregation rate will scale with the total protein concentration m_{tot}, thus:

$$k_{\mathrm{max}} = \kappa e^{-1} \tag{6.1}$$

However, the maximum aggregation rates of the bovine insulin growth curves in figure 3.1 appear to only match this prediction at low protein concentrations, and only for the fragmentation model (when $\kappa = \sqrt{2k_{+}k_{\mathrm{f}}m_{\mathrm{tot}}}$) [(1)]. This can be seen from figure 6.1.

(1) The fragmentation model predicts $k_{\mathrm{max}} \propto m_{\mathrm{tot}}^{1/2}$, while the heterogeneous nucleation model predicts $k_{\mathrm{max}} \propto m_{\mathrm{tot}}^{3/2}$ with a secondary nucleus size of $n_{\mathrm{s}} = 2$.

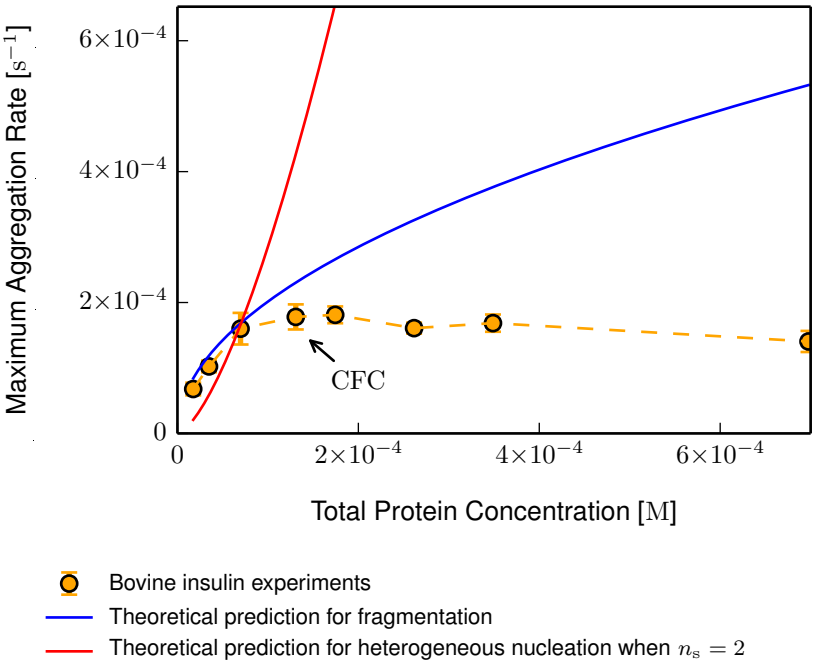


Figure 6.1 The scaling behaviour of $k_{\max}$ with m_{tot} in bovine insulin experiments, compared with the theoretical predictions (mean-field equation 6.1) of the fragmentation (with $k_f = 3\times10^{-8}\ \text{s}^{-1}$) and heterogeneous nucleation (where $k_h = 24\ \text{M}^{-1}\ \text{s}^{-1}$ and $n_s 2$) models, where $k_+ = 5\times10^{4}\ \text{M}^{-1}\ \text{s}^{-1}$ is common to both. Each point represents the mean maximum aggregation rate of between 140 and 200 replicate experiments at a particular concentration with error bars indicating the standard deviation. The arrow indicates the critical fibril concentration (CFC), which will be the focus of this chapter. These were extracted from growth curves using the method described in §2.3.2.

For the fragmentation model, while there is good agreement at low protein concentrations, the maximum aggregation rate for the experimental data saturates and possibly even decreases at higher protein concentrations. This effect cannot be explained by the experimental data having reached the diffusion limit, as the elongation rate constant $k_+ = 5\times10^{4}\ \text{M}^{-1}\ \text{s}^{-1}$ is far lower than the theoretical diffusion limit for a molecule of the size of bovine insulin; moreover, with an increased availability of monomers at higher protein concentrations, one would expect a reduction in the average time between consecutive encounters of a filament end a free monomer, not an increase.

Interestingly, the addition of salt (NaCl) to the experimental samples changes this behaviour as shown in figure 6.2. At the lowest salt concentration the maximum

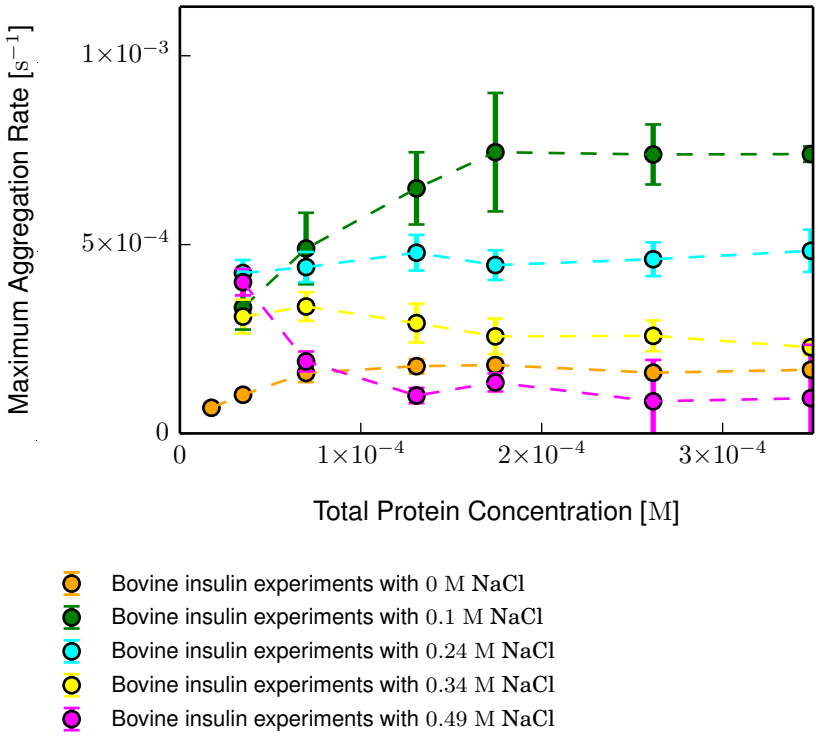


Figure 6.2 The scaling behaviour of $k_{\max}$ with m_{tot} in bovine insulin experiments when self-assembly occurs in the presence of different concentrations of NaCl salt. Each point represents the mean maximum aggregation rate of between 140 and 200 replicate experiments at a particular concentration with error bars indicating the standard deviation; the lines are to help differentiate the data sets. These maximum aggregation rates were extracted from growth curves using the method described in §2.3.2.

aggregation rate increases dramatically but the saturation behaviour is preserved. Higher concentrations of salt give a lower maximum aggregation rate, and the trend eventually inverts once the salt concentration exceeds ~ 0.4 M; i.e. for very high salt concentrations, the maximum aggregation rate decreases with increasing salt.

The lag times are also affected by the addition of salt but in a different way, as shown in figure 6.3. Here the scaling exponent γ (discussed in §5) decreases in the presence of a small concentration of salt but then increases with further addition of salt. This drop in the average lag-time is simply linked to the increase in maximum aggregation rate. As these effects are linked, both can be understood

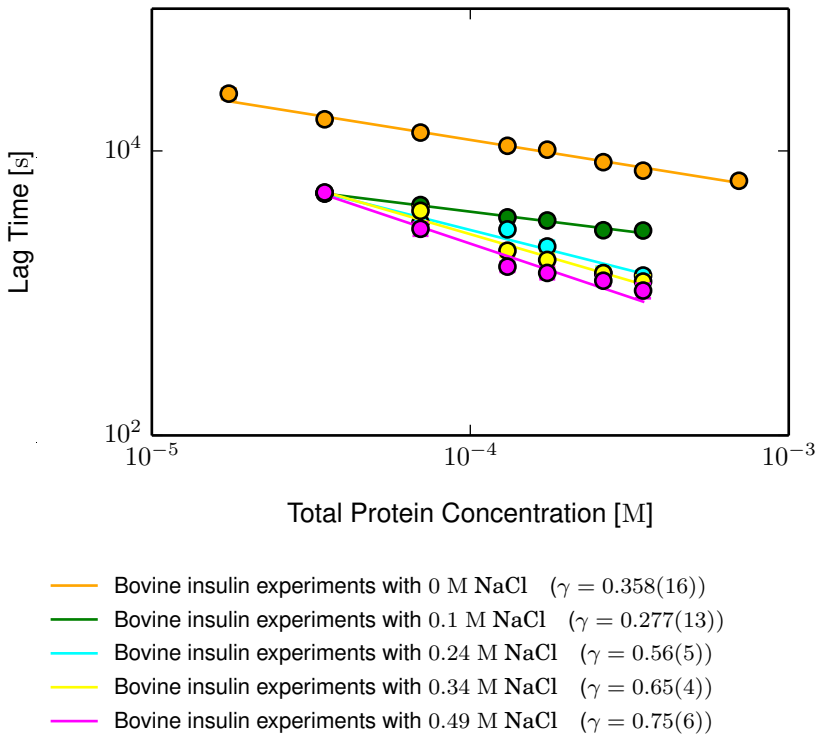


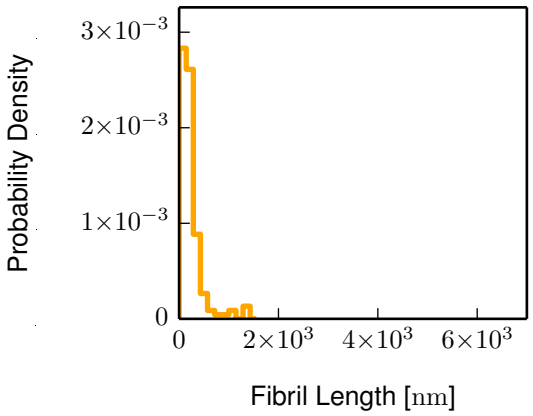
Figure 6.3 The scaling behaviour of τ_3 with m_{tot} in bovine insulin experiments when self-assembly occurs in the presence of different concentrations of NaCl salt. Each point represents the mean lag-time of between 140 and 200 replicate experiments at a particular concentration with error bars indicating the standard deviation; the lines are the best fit of the power law $\tau_{\mathrm{lag}} \propto m_{\mathrm{tot}}^{-\gamma}$ to each data set. These lag times were extracted from growth curves using the method described in §2.3.2.

by examining the filament length distributions in the absence of salt at both low and high protein concentration.

During most of the aggregation process, the fragmentation model is predicted to produce a broad, skew normal filament length distribution containing both long and short filaments. Upon completion of aggregation, this is expected to collapse into a narrow distribution containing a large number of very short filaments (see §1.3.2 and §3.3.6).

The filament length distributions obtained from TEM images of bovine insulin fibrils at low and high protein concentrations are shown in figures 6.4 and 6.5, respectively. These appear quite distinct from one another: the length distribution obtained at a low protein concentration can indeed be described as a narrow distribution, with a peak close to zero (i.e. most fibrils are very short); by contrast,

(a) Fibril length distribution.



(b) Example TEM image with a 1 μm scale bar.

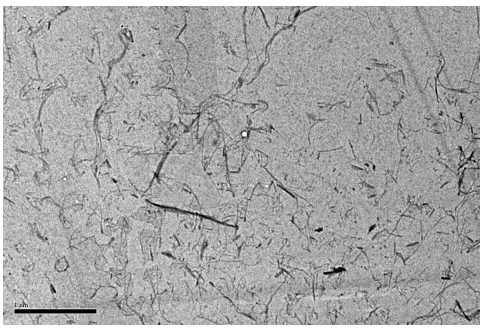
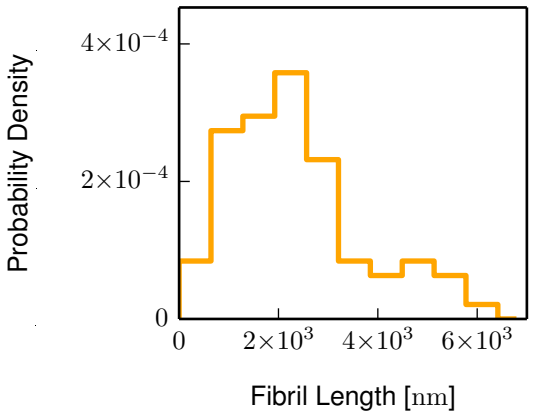


Figure 6.4 Experimental fibril length distribution **(a)**, and an example of the transmission electron microscope images from which the length distribution was obtained **(b)**, for $m_{\text{tot}} = 35$ μM bovine insulin without salt.

(a) Fibril length distribution.



(b) Example TEM image with a 0.5 μm scale bar.

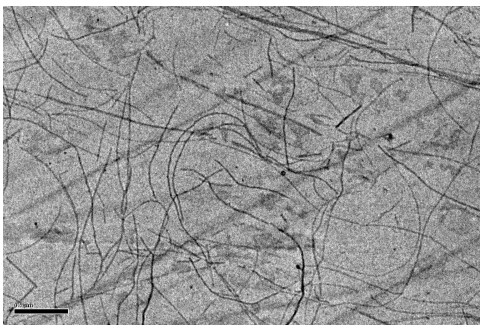


Figure 6.5 Experimental fibril length distribution **(a)**, and an example of the transmission electron microscope images from which the length distribution was obtained **(b)**, for $m_{\text{tot}} = 700$ μM bovine insulin without salt.

the length distribution obtained at a high protein concentration is much broader and clearly peaked at much longer filament lengths.

While the predicted length distribution for the fragmentation model matches the experimental data at low protein concentrations, at high protein concentrations they fail to describe both the long-time steady state length distribution and the maximum aggregation rate. Interestingly, however, the shape of the length distribution found experimentally at high protein concentrations does match the length distribution predicted by the fragmentation model at earlier times; i.e. a broad skew normal distribution, as described in §3.3.6. This observation will play a key part in the development of a model that can explain the experimental data, as I demonstrate later in this chapter.

The total protein concentration at which the maximum aggregation rate stops increasing, appears to be the same concentration at which the length distribution in the long-time limit changes shape. This value (140 μM insulin) also coincides with reported changes in the viscosity of the sample, and its circular dichroism (**CD**) spectrum, after aggregation has completed, suggesting a difference in the morphology of the fibrils [74]. The coincidence of an apparent morphological differences in the fibrils, and difference in the kinetics and fibril length distributions, all point to mechanistic changes that prevent the maximum aggregation rate from increasing beyond a certain value and prevent the length distribution from relaxing to the expected shape in the long-time limit. This change in the protein aggregation mechanism is likely to be linked to changes in the fibril structure.

In the fragmentation model, the two factors that affect the maximum aggregation rate, apart from total protein concentration, are the elongation and fragmentation rates. The shape of the experimental length distributions at the end-point, suggest that fragmentation is being suppressed at higher protein concentrations, as this would prevent the length distribution from collapsing. In the case where no salt was added, such a reduction in the fragmentation rate must happen late enough in the kinetics that it does not affect the lag time, but early enough so as to affect the maximum aggregation rate and before monomer depletion effects result in the collapse of the length distribution. Conversely, for higher salt concentrations the suppression of fragmentation must be earlier, since in this case an effect on the lag-time scaling is also observed.

An *ad hoc* way to satisfy these constraints in the model is to decrease the value of the fragmentation rate constant when a threshold concentration of fibrillar protein, or number of fibrils, has formed. This is motivated by the experimental observation that viscosity increases at higher protein concentrations, indicating the possibility that a gel network has formed which might stabilise the fibrils against breakage. The CD spectra also indicate changes in the fibril structure which might imply lateral fibril association [(2)], or "bundling", which has been observed to occur in bovine insulin [49] and could be expected to stabilise fibrils against breakage. Both of these effects would require a sufficient density of fibrils, but it is not obvious whether this pertains to the number of fibrils, or their total protein content.

Using the same kinetic Monte Carlo simulation approach as in §3, I will now explore the consequences of suppressing fragmentation: first by assuming that this happens at a critical concentration of fibrillar protein ($\mathbf{CFC}_\mathrm{M}$), and then later see how the effect changes for a critical number of fibrils ($\mathbf{CFC}_\mathrm{N}$).

6.2 A critical fibrillar protein concentration transition

In order to suppress fragmentation when the fibrillar protein concentration reaches a critical value, I first need to determine that value (the CFC_M) from the experimental data. Obviously, reducing the rate of fragmentation will only affect the maximum aggregation rate (in the way seen in figure 6.1), if the change occurs before the maximum aggregation rate has been reached; this point is typically close to $M = 0.5m_\mathrm{tot}$ in both the experimental and simulated growth curves. From figure 6.1, the maximum aggregation rate departs from the prediction when the total protein concentration is around $m_\mathrm{tot} = 140$ μM. Hence, this must be approximately double the value of the CFC_M in order for the effect to be observed at the appropriate point. For this reason I shall use $CFC_M = 70$ μM throughout this chapter unless otherwise specified.

(2) I use the term lateral association to differentiate this from filament end-joining seen in the next chapter.

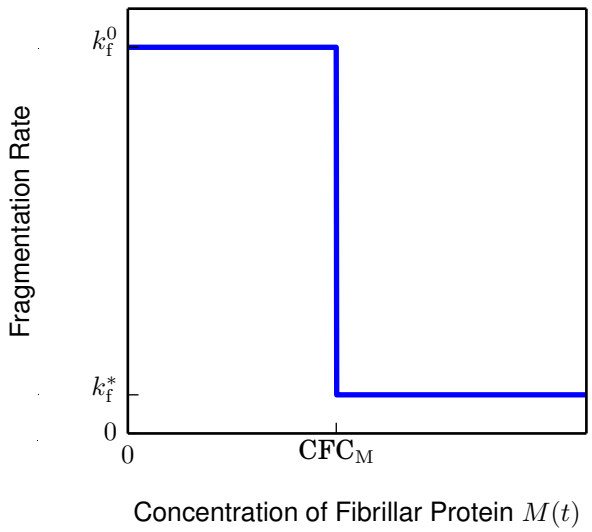


Figure 6.6 The fragmentation rate at different concentrations of fibrillar protein $M(t)$, when treated as a step transition. The fragmentation rate k_f changes abruptly once $M(t) = \mathrm{CFC_M}$, from k_f^0 when $M(t) < \mathrm{CFC_M}$, to k_f^* when $M(t) \geq \mathrm{CFC_M}$.

My first approximation for a transition in the rate of fragmentation will take the form of a step function as shown in figure 6.6. This illustrates how the fragmentation rate depends on the fibrillar protein concentration $M(t)$. Once $M(t)$ reaches the $\mathrm{CFC_M}$ the fragmentation rate changes from k_f^0 above the $\mathrm{CFC_M}$, to k_f^* below the $\mathrm{CFC_M}$, where k_f^* may be zero. This transition was added to the fragmentation model of autocatalytic self-assembly, previously discussed in §3, which included fragmentation and polymerisation; in this chapter, I do not include primary nucleation since its effects are not relevant to the phenomena under study here.

6.2.1 The effects to aspects of the growth curve, of arresting fragmentation

First, I will examine the effect of this fragmentation cutoff on the growth curves produced by kinetic Monte Carlo simulation: example growth curves are given in figure 6.7 for total protein concentrations below and above the $\mathrm{CFC_M}$. These illustrate how the growth curves of protein concentrations below the threshold are unaffected, while those above the threshold approach their plateau more gradually

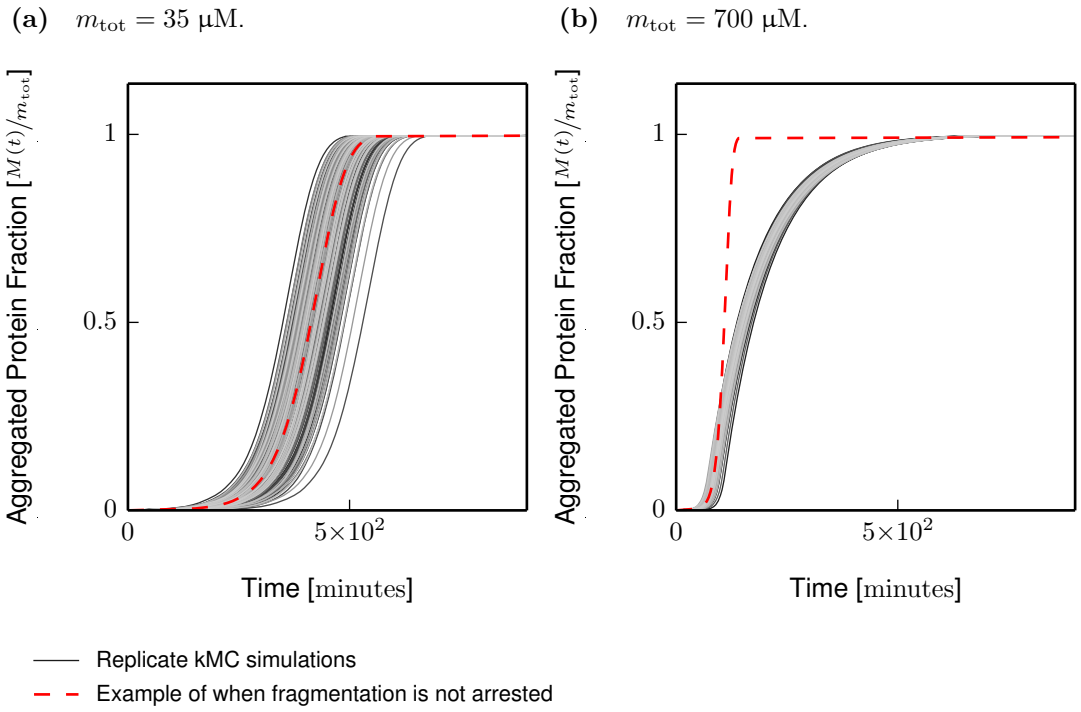


Figure 6.7 Growth curves produced by kMC simulations of a fragmentation model with fragmentation arrested once the concentration of aggregated protein reaches $\mathrm{CFC_M} = 70$ µM (i.e. $k_{\mathrm{f}}^* = 0$ in figure 6.6), for a total protein concentration: **(a)** which is not affected by the transition, and **(b)** which is affected by the transition. For both protein concentrations, 150 replicate growth curves are shown, along with an example of the growth curve produced when fragmentation is not suppressed. The common parameter values were $n_{\mathrm{c}} = 2$, $N(0) = 2 \times 10^{-12}$ M, $k_{\mathrm{n}} = 0$, $k_+ = 5{\times}10^4$ $\mathrm{M^{-1}\ s^{-1}}$, $k_- = 0$, $k_{\mathrm{f}} = 3{\times}10^{-8}$ $\mathrm{s^{-1}}$, and $V = 83{\times}10^{-14}$ l.

than when fragmentation was not suppressed, thereby shifting the maximum aggregation rate to an earlier point in the growth curve.

By definition, the lag time τ_{ϕ_M} to reach a threshold concentration of aggregated protein $M_\phi = \phi_M\, m_{\mathrm{tot}}$ (see §1.3.4) will only be affected by the arrest of fragmentation if $M_\phi \geq \mathrm{CFC_M}$. This is borne out in figure 6.8, where the only lag times affected are those for protein concentrations significantly above the $\mathrm{CFC_M}$.

The effect on the maximum aggregation rate, of arresting fragmentation in this way, can be seen in figure 6.9 for two different values of $\mathrm{CFC_M}$. This illustrates an unexpected and subtle consequence of the transition in fragmentation rate at the $\mathrm{CFC_M}$: when the value of $\mathrm{CFC_M}$ is lowered, the trend in the maximum

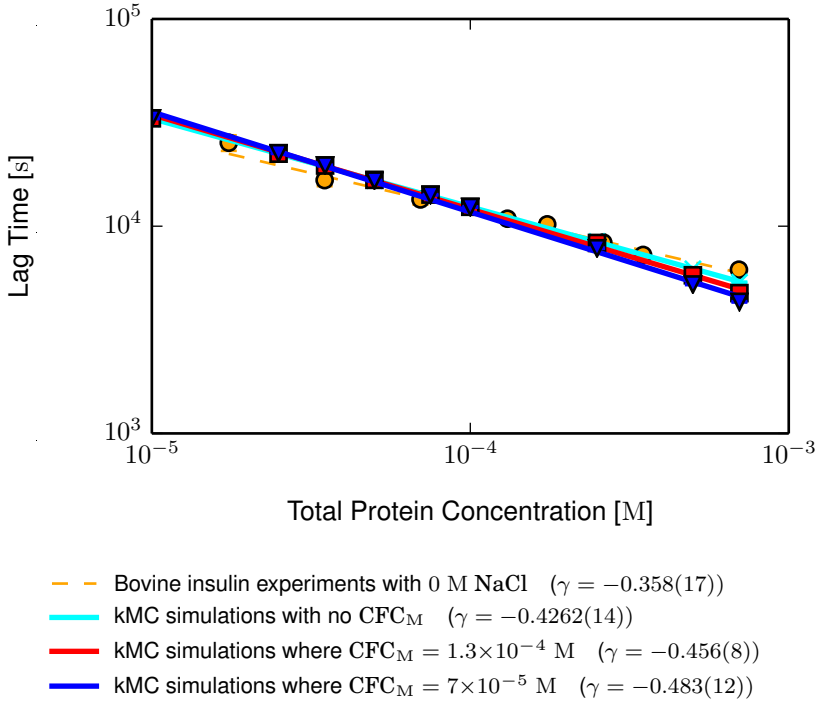


Figure 6.8 The scaling behaviour of τ_3 with m_{tot} in a fragmentation model with fragmentation arrested at different concentrations of aggregated protein, CFC_M; where the common parameter values were $n_\text{c} = 2$, $N(0) = 2 \times 10^{-12}$ M, $k_\text{n} = 0$, $k_+ = 5{\times}10^4$ M^{-1} s^{-1}, $k_- = 0$, $k_\text{f} = 3{\times}10^{-8}$ s^{-1}, and $V = 83{\times}10^{-14}$ l. Each kMC simulation point is the mean of 150 replicates with error bars indicating the standard deviation (method in §2.2). The experimental results for bovine insulin are shown for comparison. Each line indicates the best fit of the power law $\tau_\text{lag} \propto m_\text{tot}^{-\gamma}$ to the data set (method in §2.3.2).

aggregation rate changes for protein concentrations above the threshold; i.e. the curvature of the graph in figure 6.9 for the points above the CFC_M appears to invert. This effect will be important in the next section, where I examine the behaviour of the maximum aggregation rates produced experimentally in the presence of salt (shown in figure 6.3).

Thus, an abrupt transition in the rate of fragmentation upon reaching a specific concentration of aggregated material, is able to reproduce the experimentally observed effect on the maximum aggregation rate while leaving the lag times almost unaffected.

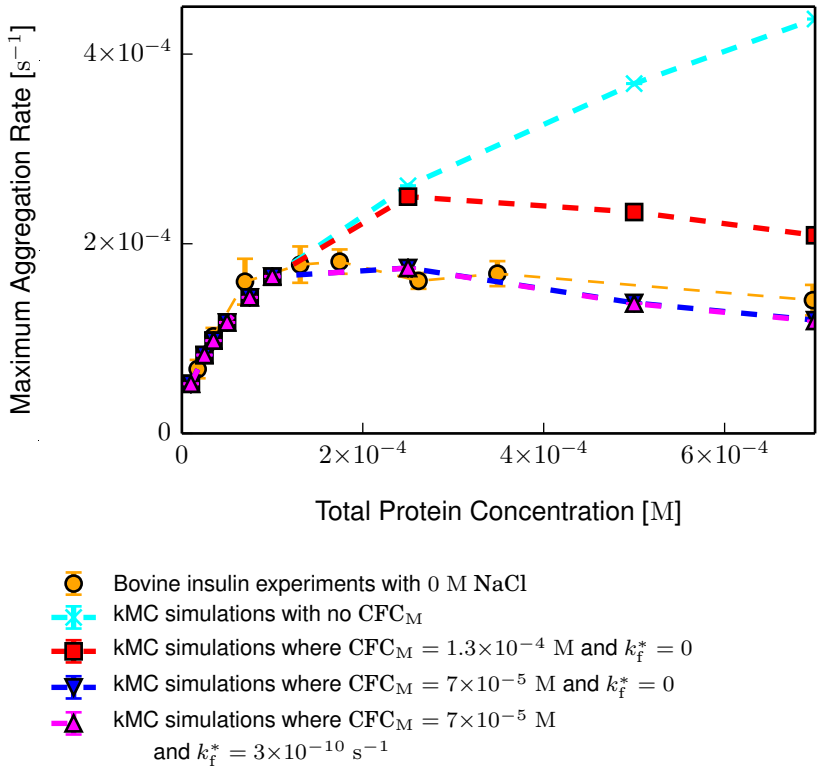


Figure 6.9 The scaling behaviour of k_{max} with m_{tot} in a fragmentation model with fragmentation arrested at different concentrations of aggregated protein, CFC_M; where the common parameter values were $n_c = 2$, $N(0) = 2 \times 10^{-12}$ M, $k_n = 0$, $k_+ = 5{\times}10^4$ M^{-1} s^{-1}, $k_- = 0$, $k_f = 3{\times}10^{-8}$ s^{-1}, and $V = 83{\times}10^{-14}$ l. Each kMC simulation point is the mean of 150 replicates with error bars indicating the standard deviation (method in §2.2); the lines are to help differentiate the data sets. For comparison, the experimental data for bovine insulin is also shown.

The influence of salt: Changing the CFC_M

As I demonstrated in figure 6.9, by lowering the CFC_M it is possible to alter the curvature of the maximum aggregation rate data for protein concentrations above the CFC_M. Hence, by appropriate choice of CFC_M, it is possible to reproduce the effect illustrated in figure 6.2, where increasing amounts of salt invert the curvature of the graphs of maximum aggregation rate plotted as a function of protein concentration. However, first I will address the large increase in maximum aggregation rate apparent for the smallest concentration of salt.

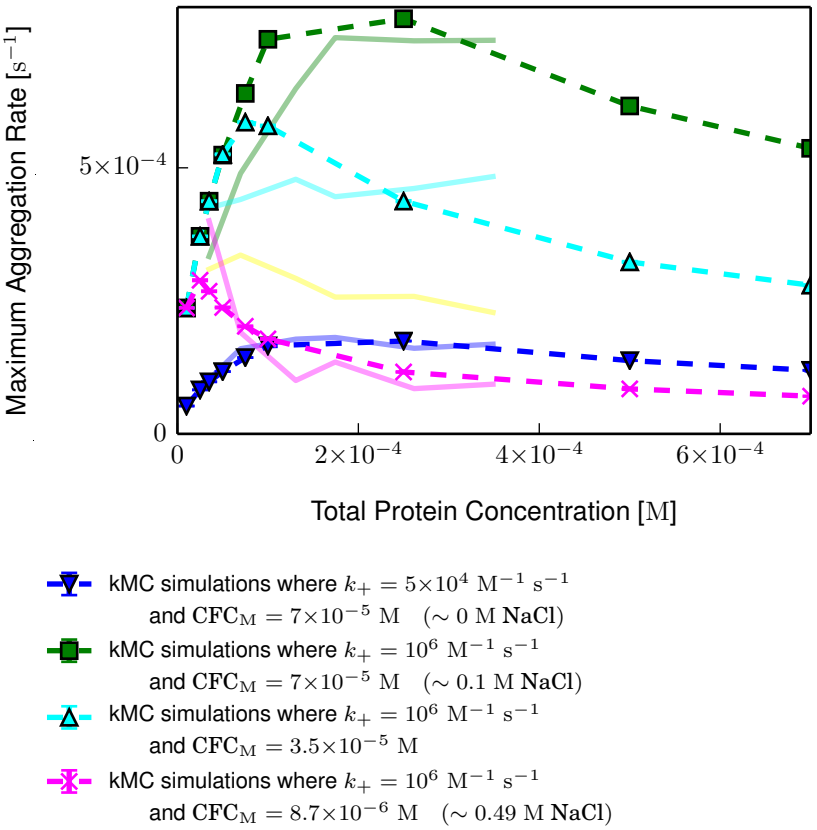


Figure 6.10 The scaling behaviour of $k_{\max}$ with m_{tot} in a fragmentation model with fragmentation arrested at different concentrations of aggregated protein, CFC_M, and with different elongation rates, k_+; where the common parameter values were $n_\text{c} = 2$, $N(0) = 2 \times 10^{-12}$ M, $k_\text{n} = 0$, $k_- = 0$, $k_\text{f} = 3\times10^{-8}$ s^{-1}, and $V = 83\times10^{-14}$ l. Each kMC simulation point is the mean of 150 replicates with error bars indicating the standard deviation (method in §2.2); the lines are to help differentiate the data sets. The bovine insulin experimental data in figure 6.2, are also shown here as **pale solid lines**, for qualitative comparison.

It is expected that the presence of salt will lower the Coulomb repulsion between free monomers and growing fibril ends, by screening the charge on both species. This will effectively increase the rate of elongation and hence increase the maximum aggregation rate. The extent to which it does this is unknown, but it is possible to match the experimental data for the lowest salt concentration in figure 6.2, by increasing the elongation rate of the simulation by a factor of 20. The effect of doing this is illustrated in figure 6.10, alongside the experimental data from figure 6.2.

Above a certain salt concentration, the charge screening effect is expected to saturate such that the elongation rate will not increase further by the addition of more salt. For the purposes of this model, I will assume that this occurs at around 0.1 M NaCl, since this is where the highest values of the maximum aggregation rate are seen in the experimental data. The decrease in the maximum aggregation rate upon further addition of salt can instead be explained by reducing the value of the CFC_M, as is also illustrated in figure 6.10 (where the higher elongation rate has been kept).

This decrease of the CFC_M value with increasing salt concentration can also be rationalised by considering the effects of charge screening. The CFC_M, as I have defined it, represents a condition at which the mechanical properties of the filaments change in conjunction with some sort of lateral filament association [(3)]. Much like monomer addition, this lateral filament association is likely to be hindered by Coulomb repulsion and thus the addition of salt may increase the rate at which filaments associate, in much the same way as it increases the rate of elongation. Since I do not consider the rate of lateral association explicitly in the simulations, this effect can be manifested as a decrease in the CFC_M value in this model.

The surface area of a filament is significantly larger than that of a growth-competent end, and so I expect screening effect to saturate more rapidly for the elongation rate than for lateral filament association. Consequently, even though the elongation rate quickly saturates, the addition of more salt could be found to lower the CFC_M further, below the minimum value (corresponding to the highest salt concentration) shown in figure 6.10.

(3) The change in sample viscosity and CD spectrum are both assumed to indicate structural differences between fibrils formed at protein concentrations above and below the CFC_M.

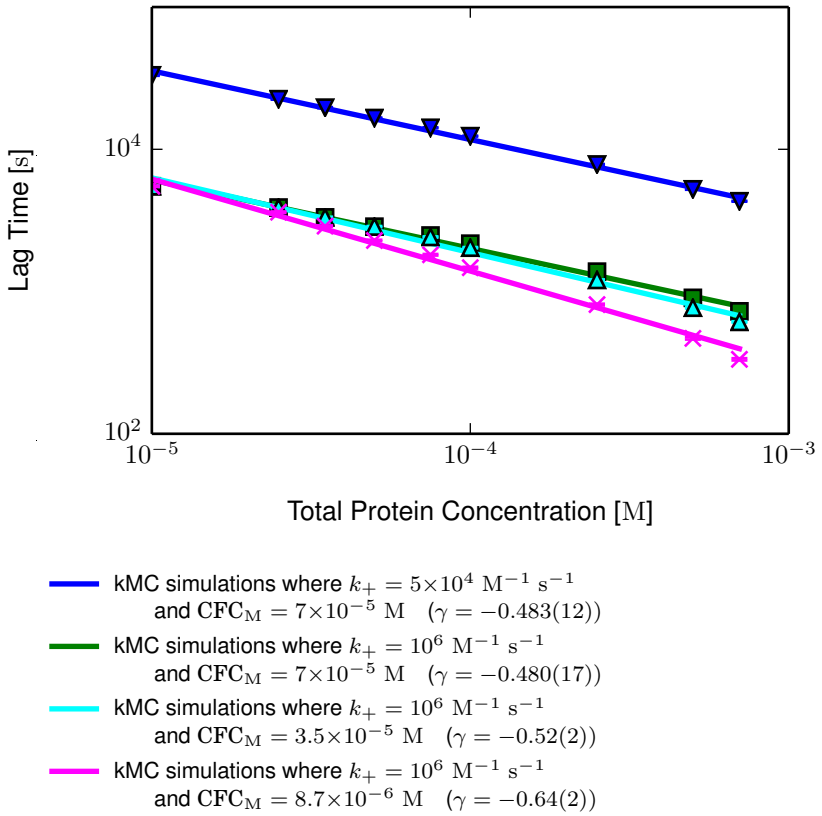


Figure 6.11 The scaling behaviour of τ_3 with m_tot in a fragmentation model with fragmentation arrested at different concentrations of aggregated protein, $\mathrm{CFC_M}$, and with different elongation rates, k_+; where the common parameter values were $n_\mathrm{c} = 2$, $N(0) = 2 \times 10^{-12}$ M, $k_\mathrm{n} = 0$, $k_- = 0$, $k_\mathrm{f} = 3{\times}10^{-8}$ s^{-1}, and $V = 83{\times}10^{-14}$ l. Each kMC simulation point is the mean of 150 replicates with error bars indicating the standard deviation (method in §2.2). Each line indicates the best fit of the power law $\tau_\mathrm{lag} \propto m_\mathrm{tot}^{-\gamma}$ to the data set (method in §2.3.2). The bovine insulin experimental data in figure 6.3 has not been shown here because the trends are too similar and the figure would become unclear.

It can be seen that this combination of raising the elongation rate and then lowering the CFC_M, qualitatively reproduces the effects of salt on the maximum aggregation rate as observed experimentally (figure 6.10). In addition it also captures the effect that the addition of salt has on the lag time, by similarly lowering the lag time and then increasing the concentration dependence, as shown in figure 6.11.

6.2.2 The effects to the filament length distribution, of arresting fragmentation

Having established that the above mechanistic transition is able to reproduce the maximum aggregation rate and lag times observed experimentally, I will now look at its effects on the filament length distribution.

For total protein concentrations above the CFC_M, suppression of fragmentation can prevent the length distribution from relaxing into the shape expected in the long-time steady state. When $k_f^* > 0$, the same length distribution produced by concentrations below the CFC_M will result at very long times, but the time taken for it to reach this state will increase with decreasing k_f^*, so that the apparent end-point length distribution may be very different. Only in the limit where $k_f^* = 0$ can the shape of the length distribution strictly be frozen.

Figure 6.12 shows the end-point length distributions for the growth curves in figure 6.7. As expected, no effect is seen for the protein concentration below the CFC_M, but the length distribution obtained at a protein concentration above the CFC_M now takes the form of a skew normal distribution (i.e. the shape present at the earlier time when it became frozen). While this is in qualitative agreement with the experimental data, it is clear that this simulated length distribution is significantly narrower than the length distribution found experimentally and also has a much longer mean filament length.

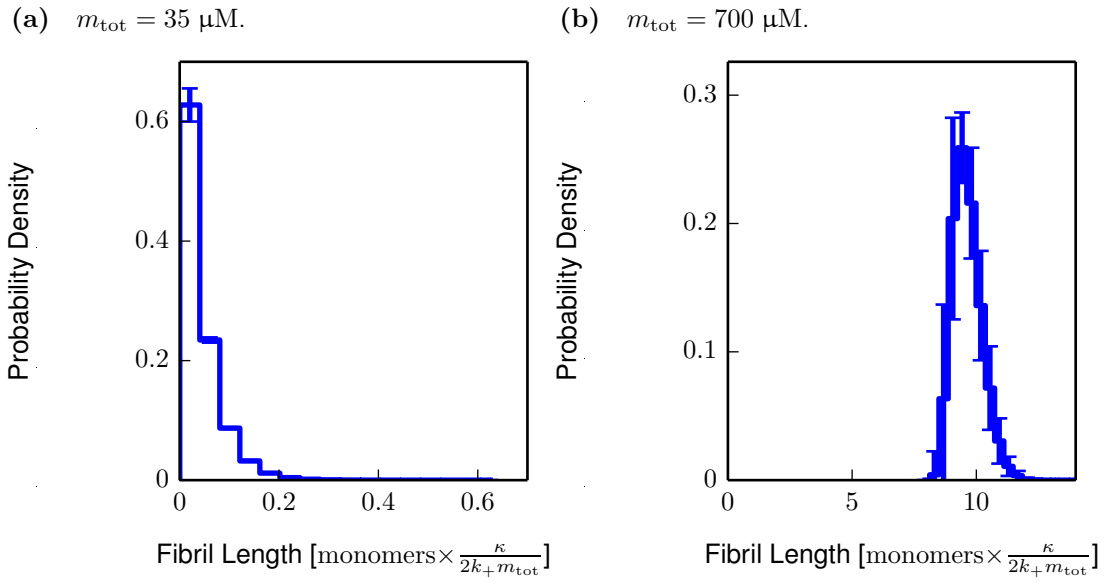


Figure 6.12 Average filament length distributions at the end-point of replicate kMC simulations of a fragmentation model with fragmentation arrested once the concentration of aggregated protein reaches $\mathrm{CFC_M} = 70$ μM (i.e. $k_{\mathrm{f}}^* = 0$ in figure 6.6), for a total protein concentration: **(a)** that is not affected by the transition, and **(b)** that is affected by the transition; where the common parameter values were $n_{\mathrm{c}} = 2$, $N(0) = 2 \times 10^{-12}$ M, $k_{\mathrm{n}} = 0$, $k_- = 0$, $k_{\mathrm{f}} = 3{\times}10^{-8}$ s^{-1}, and $V = 83{\times}10^{-14}$ l. Filament lengths have been normalised by the predicted mean filament length $2k_+ m_{\mathrm{tot}}/\kappa$ (equation 6.3). Each distribution represents the mean of 150 replicate simulations with error bars indicating the standard deviation (determined using the method in §2.2).

I have presented the distributions in figure 6.12 in terms of the mean filament length at equilibrium so as to remove the dependence on the choice of elongation and fragmentation rate (as discussed previously in §3.3.6). By doing this, I also highlight the large difference between the mean filament length of these two distributions. The origin of this difference can be understood by considering again the mechanisms which lead to the formation of the length distribution; i.e. that longer filaments are more likely to fragment, thus driving the length distribution toward shorter filaments, while elongation acts to push the distribution towards longer filaments.

The mean filament length for protein concentrations above the CFC

When fragmentation is arrested, the length distribution will be frozen but some monomer may remain, not yet having been incorporated into filaments. This extra material has an equal probability of being added to any filament. Thus, the shape of the length distribution is expected to remain unchanged but the average length of each filament is expected to increase according to the number of filaments and free monomers present. The extent of this shift in the mean filament length can be expressed analytically as follows.

The mean filament length at time t is given by the ratio of the fibrillar protein content $M(t)$ and the number of fibrils $N(t)$ (equation 1.23 in §1.3.4):

$$\overline{L}(t) = \frac{M(t)}{N(t)} \tag{6.2}$$

where the expressions $M(t)$ and $N(t)$ depend on the rates of elongation and fragmentation up to that time. For the steady-state condition describing the length distribution at intermediate times — i.e. when the system has relaxed and the monomer concentration can be assumed constant [13] — this gives the approximate mean filament length:

$$\overline{L}(t) \approx \frac{\kappa}{k_{\mathrm{f}}} \tag{6.3}$$

If the length distribution is frozen, the remaining free monomer will be

$$m(t_{\mathrm{CFC_M}}) = m_{\mathrm{tot}} - M(t_{\mathrm{CFC_M}})$$

which, if added uniformly to all filaments, will cause the mean (given by equation 6.3) to increase by approximately

$$\frac{m_{\mathrm{tot}} - M(t_{\mathrm{CFC_M}})}{N(t_{\mathrm{CFC_M}})}$$

From this, the mean filament length of the end-point length distribution can be found analytically in the case where $k_{\mathrm{f}}^{*} = 0$, by equating equations 6.2 and 6.3, giving

$$N(t_{\mathrm{End}}) = N(t_{\mathrm{CFC_M}}) \simeq \frac{k_{\mathrm{f}}}{\kappa}\mathrm{CFC_M}$$

which defines the extent to which equation 6.3 is increased, and gives in the actual mean filament length at the end point

$$\overline{L}(t_{\text{end}}) = \overline{L}(t_{\text{equilibrium}})\frac{m_{\text{tot}}}{\text{CFC}_{\text{M}}} \simeq \frac{\kappa}{k_{\text{f}}}\frac{m_{\text{tot}}}{\text{CFC}_{\text{M}}} \tag{6.4}$$

Thus, for a concentration of 700 μM with $\text{CFC}_{\text{M}} = 70$ μM, equation 6.4 correctly predicts that the mean of the filament length distribution will be an order of magnitude higher than given by equation 6.3 (as apparent in figure 6.12b).

From this it can be seen that, in order for the simulation to reproduce the experimental length distribution at high protein concentrations, it would be necessary for more filaments to be present at the point where fragmentation is arrested. As I will now discuss, this cannot be accomplished simply by increasing the fragmentation rate prior to the threshold.

6.2.3 Softening the transition

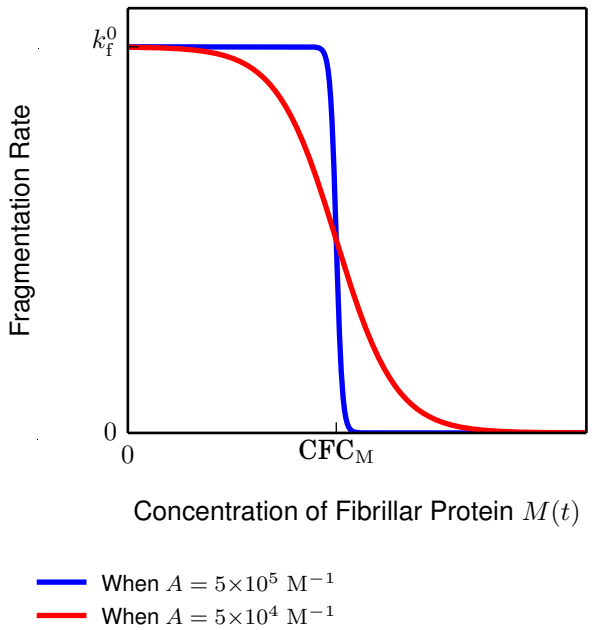


Figure 6.13 The fragmentation rate at different concentrations of fibrillar protein $M(t)$, when treated as a smooth transition, for two value of the "sharpness" constant A. The fragmentation rate k_{f} changes gradually as $M(t)$ crosses CFC_{M}, from k_{f}^0 when $M(t) < \text{CFC}_{\text{M}}$, and decaying to zero.

The simplicity of the sharp transition is appealing, particularly as it reproduces most of the behaviour seen experimentally. However, the physical processes which this transition is approximating are unlikely to be so abrupt. Lateral association of filaments is likely to be a much more gradual process which begins while $M(t)$ is still below the $\mathrm{CFC_M}$, and continues even after it has been passed. For this reason, I will now discuss a smoothed step transition described by the function

$$k_\mathrm{f}(M) = \frac{k_\mathrm{f}^0}{2}\big(1 - \tanh\left(A[M(t) - \mathrm{CFC_M}]\right)\big)$$

where A is the transition sharpness, and it has been assumed that $k_\mathrm{f}^* = 0$. The shape of this transition is illustrated in figure 6.13.

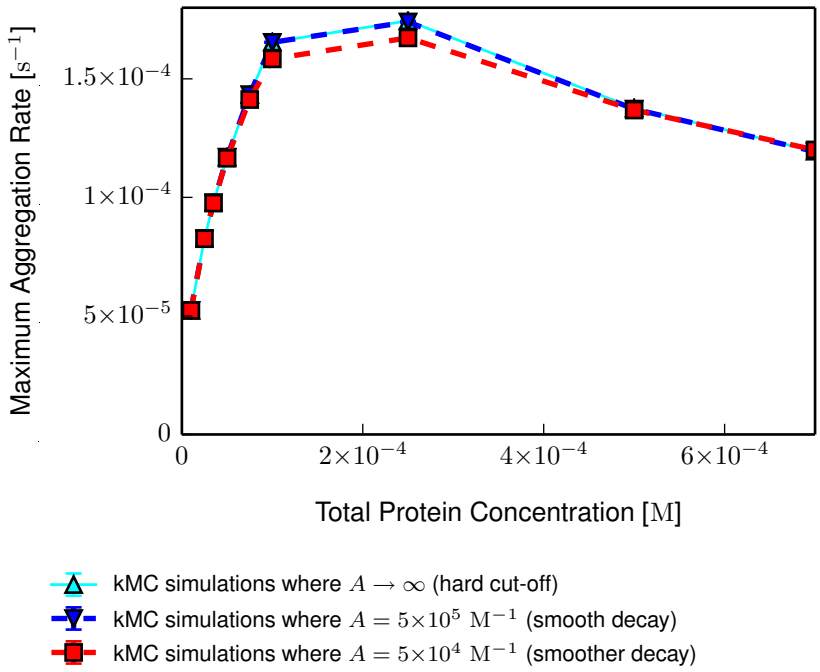


Figure 6.14 The scaling behaviour of k_max with m_tot in a fragmentation model with fragmentation gradually suppressed, according to the factor A, as the concentration of aggregated protein increases; where the common parameter values were $\mathrm{CFC_M} = 70\ \mu\mathrm{M}$, $n_\mathrm{c} = 2$, $N(0) = 2 \times 10^{-12}$ M, $k_\mathrm{n} = 0$, $k_+ = 5{\times}10^4\ \mathrm{M^{-1}\,s^{-1}}$, $k_- = 0$, $k_\mathrm{f} = 3{\times}10^{-8}\ \mathrm{s^{-1}}$, and $V = 83{\times}10^{-14}$ l. Each kMC simulation point is the mean of 150 replicates with error bars indicating the standard deviation (method in §2.2); the lines are to help differentiate the data sets.

For this gradual cutoff, the growth curves and lag times are affected only as much as they were by the abrupt cutoff; and only when the transition is very soft, is

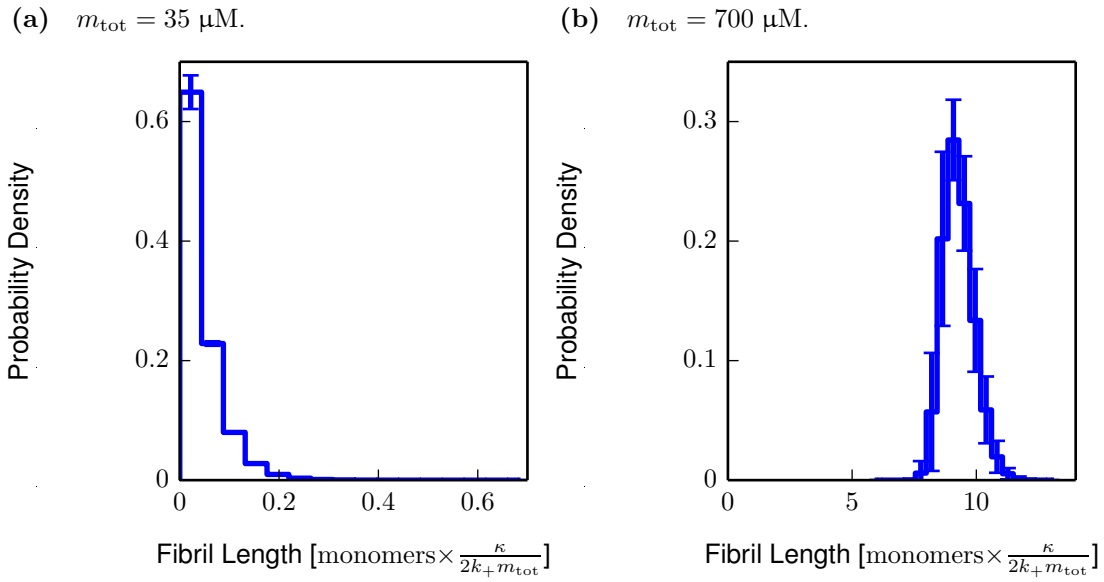


Figure 6.15 Average filament length distributions at the end-point of replicate kMC simulations of a fragmentation model with fragmentation gradually suppressed (as in figure 6.13) as the concentration of aggregated protein increases, for a total protein concentration: **(a)** that is not affected by the transition, and **(b)** that is affected by the transition; where the common parameter values were $A = 5{\times}10^4$ M^{-1},CFC$_{\mathrm{M}} = 70$ μM, $n_{\mathrm{c}} = 2$, $N(0) = 2 \times 10^{-12}$ M, $k_{\mathrm{n}} = 0$, $k_- = 0$, $k_{\mathrm{f}} = 3{\times}10^{-8}$ s^{-1}, and $V = 83{\times}10^{-14}$ l. Filament lengths have been normalised by the predicted mean filament length $2k_+ m_{\mathrm{tot}}/\kappa$ (equation 6.3). Each distribution represents the mean of 150 replicate simulations with error bars indicating the standard deviation (determined using the method in §2.2).

an effect apparent in the maximum aggregation rate, as shown in figure 6.14. However, the end-point length distributions shown in figure 6.15 are also largely unaffected by this smoothed cutoff: for the protein concentration below the CFC$_{\mathrm{M}}$ there is no difference, while the protein concentration above the CFC$_{\mathrm{M}}$ only displays a slight broadening and shift of the mean towards shorter filament lengths. Hence, this does not constitute a significant improvement over the simpler model, and similar effects on the kinetics and the length distributions are also found for other, more complicated forms of cutoff function.

The reason why changing the cutoff function makes little difference compared to the simple, abrupt transition is that the number of filaments present, simultaneously defines the length distribution, the maximum aggregation rate,

and other aspects of the kinetics. This number is largely formed during a narrow window of time at the end of the lag phase. Consequently a smoothly decaying fragmentation rate for concentrations above the $\mathrm{CFC_M}$, will have little effect as most of the filament population is formed before the $\mathrm{CFC_M}$ is reached; and while the contribution to the kinetics after the $\mathrm{CFC_M}$ will be important for keeping the mean filament length down, it also raises the maximum aggregation rate.

In conclusion, if the CFC is defined in terms of a threshold concentration of fibrillar protein — and almost regardless of how the transition is implemented — the reduction in fragmentation rate upon crossing the threshold gives a satisfactory explanation of the lag time and maximum aggregation rate trends. It also explains the difference in shape of the length distributions for protein concentration below and above the $\mathrm{CFC_M}$, where those above the threshold are frozen in their early-time state but shifted towards longer filaments lengths.

I will now look at the effects of defining the CFC in terms of the number concentration of filaments $\mathrm{CFC_N}$ rather than $\mathrm{CFC_M}$ and in doing so, resolve the issue with the length distributions.

6.3 A critical fibril number concentration transition

Defining the critical fibril concentration (CFC) as a quantity of protein incorporated into fibrils was the obvious choice since the value of the $\mathrm{CFC_M}$ could easily be deduced from the experimental kinetics. This is not so intuitive for a transition in fragmentation rate that occurs at a critical fibril number concentration ($\mathrm{CFC_N}$).

So as to produce an effect comparable with that of the transition at the $\mathrm{CFC_M}$, I will define $\mathrm{CFC_N}$ as the number of filaments present in the equilibrium length distribution for a protein concentration $m_{\mathrm{tot}} = 2\mathrm{CFC_M}$. I can then use the definition of the mean filament length given in equations 6.2 and 6.3 to write

$$\frac{2\mathrm{CFC_M}}{\mathrm{CFC_N}} = \frac{\kappa}{k_{\mathrm{f}}}$$

where $\kappa = \sqrt{2k_+k_\mathrm{f}m_\mathrm{tot}}$, from which the relation between $\mathrm{CFC_N}$ and $\mathrm{CFC_M}$ can be obtained in terms of the rate constants:

$$\mathrm{CFC_N} = \sqrt{\frac{2\mathrm{CFC_M}k_\mathrm{f}}{k_+}} \tag{6.5}$$

It is worth noting that for a threshold defined in terms of the number of filaments, any total protein concentration can reach the threshold provided ${}^{m_\mathrm{tot}}/_{n_\mathrm{c}} > \mathrm{CFC_M} = 70\ \mu\mathrm{M}$. This is because fragmentation will continue to generate new filaments until either the threshold is reached or all filaments are reduced to their minimum stable length. In practice however, the protein concentration at which this occurs will be higher than I have indicated due to the fact that even the shortest stable filament can fragment, thereby reducing the number of filaments. Consequently, some of the protein concentrations which were unaffected in the previous section by the $\mathrm{CFC_M}$, may now be affected by the $\mathrm{CFC_N}$. Additionally, low total protein concentrations will be arrested with a proportionally higher fraction of the monomer population in filaments, than higher concentrations [(4)]. This will have important ramifications for the length distribution, discussed below.

For simplicity, I will reuse the step transition function from §6.2 (illustrated for the $\mathrm{CFC_N}$ in figure 6.16). This I implemented in kinetic Monte Carlo simulations of the fragmentation model as before, with the fragmentation rate going to zero once $N(t) = \mathrm{CFC_N}$.

Effects on aspects of the growth curve

The point in the kinetics where the $\mathrm{CFC_N}$ is reached, depends on the total protein concentration: when $m_\mathrm{tot} > \mathrm{CFC_M}$ (used in equation 6.5 to define $\mathrm{CFC_N}$), the $\mathrm{CFC_N}$ will be reached after passing the $\mathrm{CFC_M}$; when $m_\mathrm{tot} < \mathrm{CFC_M}$, the $\mathrm{CFC_N}$ must necessarily be reached first (since, by definition, the $\mathrm{CFC_M}$ can never be reached). However, for all total protein concentrations considered, the maximum aggregation rate is always reached before the critical number of fibrils have formed. Consequently, the change in fragmentation rate at the transition has

(4) Low protein concentrations will have fewer filaments relative to m_tot, than will higher protein concentrations.

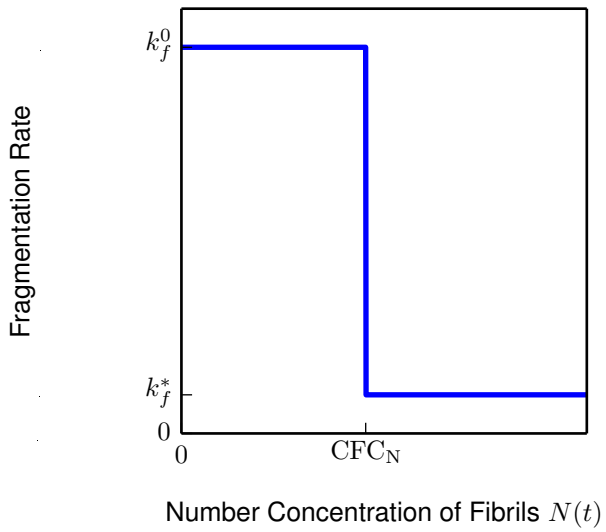


Figure 6.16 The fragmentation rate at different concentrations of fibrils $N(t)$, when treated as a step transition. The fragmentation rate k_f changes abruptly once $N(t) = CFC_N$, from k_f^0 when $N(t) < CFC_N$, to k_f^* when $N(t) \geq CFC_N$.

little effect on the lag time or maximum aggregation rate, as will be seen in figures 6.19 and 6.20, which I present shortly in §6.4.

Effects on the filament length distribution

While the lack of effect on the maximum aggregation rate indicates that this CFC_N model is not able to explain that aspect of the experimental data, the effects which it has on the length distributions are more encouraging.

Shown in figure 6.17 are the end-point length distributions produced by the model implementing a change in fragmentation rate at the CFC_N. These are similar to those shown previously in figure 6.12, where the fragmentation rate changed at the CFC_M. In figure 6.17, both the low and high protein concentrations are arrested in such a way as to produce a shape more closely resembling that of the experimental length distributions shown in figures 6.4 and 6.5. Moreover, the mean filament length in both cases has changed dramatically: the low protein concentration here has a higher mean filament length than seen in the CFC_M model, while the high protein concentration has a lower mean filament length

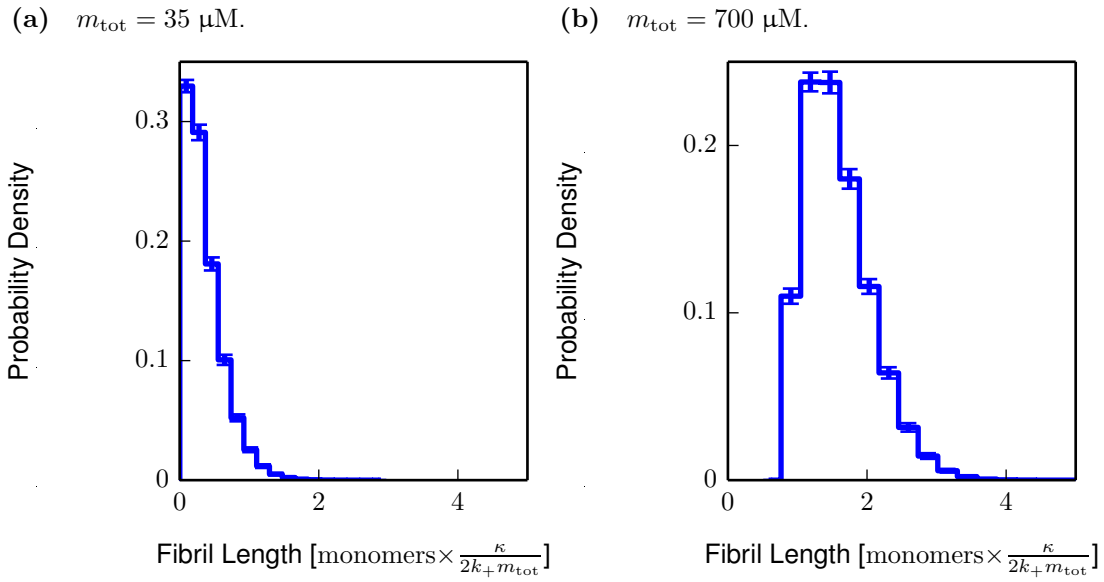


Figure 6.17 Average filament length distributions at the end-point of replicate kMC simulations of a fragmentation model with fragmentation arrested once the concentration of fibrils reaches $\mathrm{CFC_N} = \sqrt{2\mathrm{CFC_M}k_{\mathrm{f}}/k_+} = 9.2{\times}10^{-3}$ μM (i.e. $k_{\mathrm{f}}^* = 0$ in figure 6.16), for a total protein concentration: **(a)** where $m_{\mathrm{tot}} < \mathrm{CFC_M}$, and **(b)** where $m_{\mathrm{tot}} > \mathrm{CFC_M}$; where the common parameter values were $n_{\mathrm{c}} = 2$, $N(0) = 2 \times 10^{-12}$ M, $k_{\mathrm{n}} = 0$, $k_- = 0$, $k_{\mathrm{f}} = 3{\times}10^{-8}$ s^{-1}, and $V = 83{\times}10^{-14}$ l. Filament lengths have been normalised by the predicted mean filament length $2k_+ m_{\mathrm{tot}}/\kappa$ (equation 6.3). Each distribution represents the mean of 150 replicate simulations with error bars indicating the standard deviation (determined using the method in §2.2).

than seen in that model. This is due to the earlier arrest of fragmentation at lower protein concentrations.

Since the $\mathrm{CFC_N}$ occurs later in the kinetics than the $\mathrm{CFC_M}$ (for protein concentrations higher than $\mathrm{CFC_M}$) for the parameters used here, it is possible to combine their effect provided that for the $\mathrm{CFC_M}$, $k_{\mathrm{f}}^* > 0$, while for the $\mathrm{CFC_N}$, $k_{\mathrm{f}}^* = 0$. In this way fragmentation would be suppressed early enough to affect the maximum aggregation rate, but arrested late enough that a sufficient number of filaments are present to keep the mean filament length low.

6.4 Two transitions: combining the $\mathrm{CFC_M}$ and $\mathrm{CFC_N}$ models

I will now explore the effects of combining both transitions into a single model. There are a number of ways by which this could be done, but I will only consider the case where the fragmentation rate is reduced to a non-zero value k_f^* at the $\mathrm{CFC_M}$ and then to zero at the $\mathrm{CFC_N}$. The reason for this choice is obvious: only protein concentrations above the $\mathrm{CFC_M}$ can be affected by it, and so arresting fragmentation upon crossing the $\mathrm{CFC_M}$ will produce the same affects as in §6.2.

I tested several different values of the lower fragmentation rate k_f^* and found that these produce very similar behaviour to that given below for $k_\mathrm{f}^* = 0.1k_\mathrm{f}$, except when k_f^* is either very low or very high: when k_f^* is very low, the results tended towards those where fragmentation is arrested at the $\mathrm{CFC_M}$ (see §6.2), while higher values — closer to the initial fragmentation rate — produce similar results to those where only the $\mathrm{CFC_N}$ was implemented (see §6.3).

Effects on the filament length distribution

The filament length distributions produced when both transitions in the fragmentation rate are considered in the same model, are shown in figure 6.18. These can be compared with the length distributions obtained by either of the transitions individually (figures 6.12 and 6.17), revealing that this combination produces length distributions very similar in appearance to those of the $\mathrm{CFC_N}$ model, and importantly, the experimental data (also shown in figure 6.18).

For $m_\mathrm{tot} < \mathrm{CFC_M}$, the narrow distribution given in figure 6.17 is again produced because the only threshold that is crossed at these protein concentrations is the $\mathrm{CFC_N}$. For $m_\mathrm{tot} > \mathrm{CFC_M}$, however, the length distribution is much broader and contains shorter filament lengths than the length distributions produced by either of the thresholds individually.

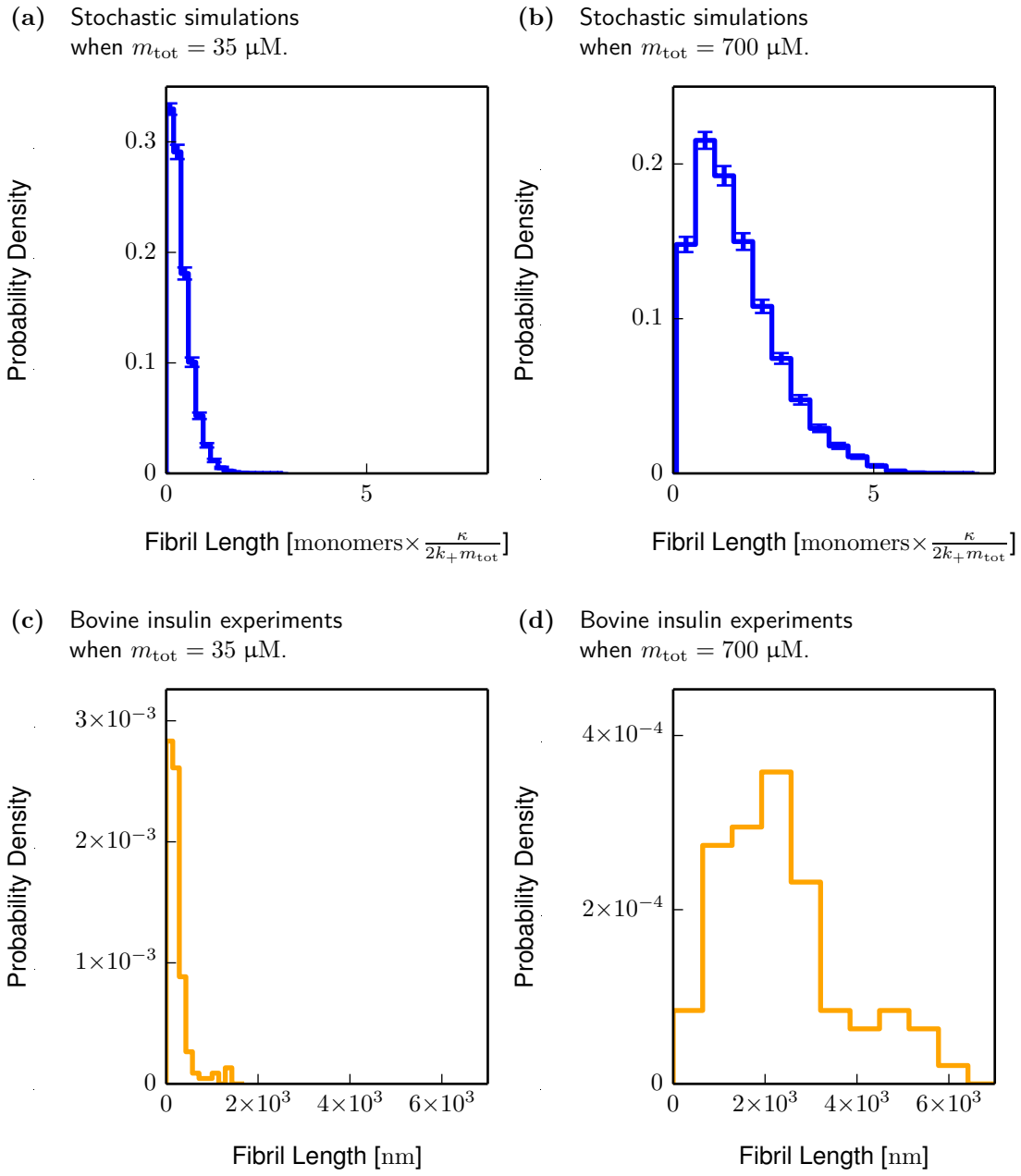


Figure 6.18 Filament length distributions from kMC simulations and bovine insulin experiments. **(a)** and **(b)** are the average, at the end-point, of replicate simulations of a fragmentation model with fragmentation slowed to $k_{\mathrm{f}}^{*} = 3{\times}10^{-9}\ \mathrm{s}^{-1}$ once $M(t) = \mathrm{CFC_M} = 70$ µM, and arrested once $N(t) = \mathrm{CFC_N} = \sqrt{2\mathrm{CFC_M}k_{\mathrm{f}}/k_{+}}$, for two total protein concentrations; where the common parameter values were $n_{\mathrm{c}} = 2$, $N(0) = 2 \times 10^{-12}$ M, $k_{\mathrm{n}} = 0$, $k_{-} = 0$, $k_{\mathrm{f}} = 3{\times}10^{-8}\ \mathrm{s}^{-1}$, and $V = 83{\times}10^{-14}$ l. Filament lengths have been normalised by the predicted mean filament length $2k_+ m_{\mathrm{tot}}/\kappa$ (equation 6.3). Each distribution represents the mean of 150 replicate simulations with error bars indicating the standard deviation (see §2.2). **(c)** and **(d)** are the experimental data previously shown in figures 6.4 and 6.5, respectively.

From this, it appears that the experimental length distributions at high protein concentrations can be explained by the suppression of fragmentation early on in the kinetics. However, in order to explain the experimental length distributions at all protein concentrations the fragmentation rate also needs to be fully arrested later in the kinetics. This has an interesting effect on the maximum aggregation rate as I will explain next.

Effects on aspects of the growth curve

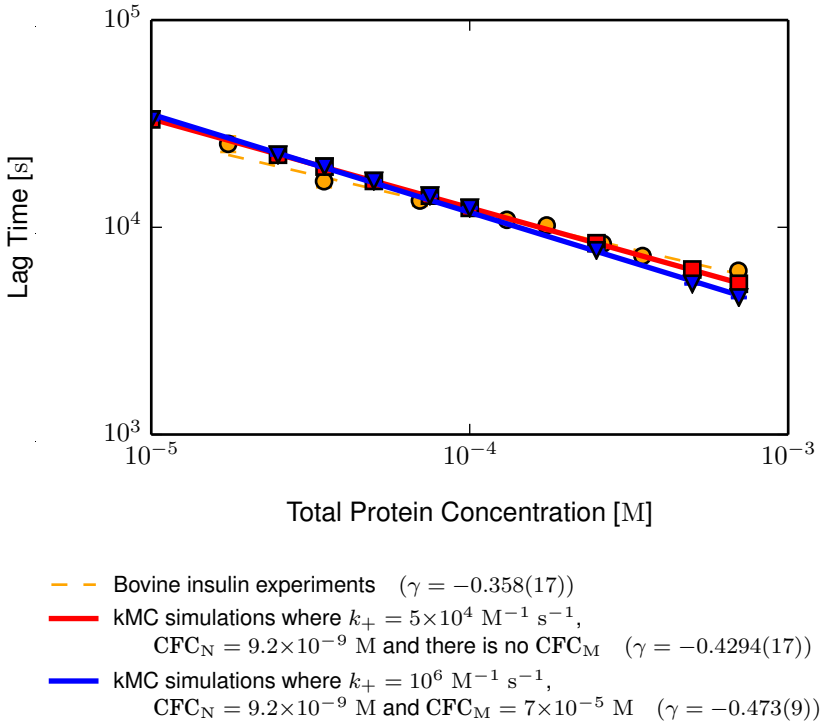


Figure 6.19 The scaling behaviour of τ_3 with m_{tot} in a fragmentation model with fragmentation slowed to $k_{\mathrm{f}}^* = 3\times10^{-9}$ s^{-1} once $M(t) = \mathrm{CFC_M}$ (if applicable), and arrested once $N(t) = \mathrm{CFC_N} = \sqrt{2\mathrm{CFC_M}k_{\mathrm{f}}/k_+}$; where the common parameter values were $n_{\mathrm{c}} = 2$, $N(0) = 2\times10^{-12}$ M, $k_{\mathrm{n}} = 0$, $k_+ = 5\times10^4$ M^{-1} s^{-1}, $k_- = 0$, $k_{\mathrm{f}} = 3\times10^{-8}$ s^{-1}, and $V = 83\times10^{-14}$ l. Each kMC simulation point is the mean of 150 replicates with error bars indicating the standard deviation (method in §2.2). The experimental results for bovine insulin are shown for comparison. Each line indicates the best fit of the power law $\tau_{\mathrm{lag}} \propto m_{\mathrm{tot}}^{-\gamma}$ to the data set (method in §2.3.2).

Since both of the threshold conditions are sufficiently late in the kinetics, they have little effect on the lag time; indeed, the only change in lag time occurs when the fragmentation rate is reduced at the CFC_M. This is evident from figure 6.19, which also includes the case where the fragmentation rate is only affected by the CFC_N.

A similar effect is also found for the maximum aggregation rate with the original elongation rate (figure 6.20). When there is a reduction in the fragmentation rate only at the CFC_N, the maximum aggregation rate scales as it did for the fragmentation model without a transition. On the other hand, if the fragmentation rate is also reduced at the CFC_M, the maximum aggregation rate behaves as it did in §6.2 without the cutoff at the CFC_N, now scaling with total protein concentration according to the reduced fragmentation rate k_f^*:

$$k_{max} = \sqrt{2k_+k_f^*m_{tot}}e^{-1}$$

If the elongation rate is increased, however, then longer filaments will be produced earlier in the kinetics. This results in an increased rate of filament formation through fragmentation, despite the rate constant having been reduced upon passing the CFC_M threshold. Consequently, more filaments are formed until the CFC_N is reached and fragmentation is arrested (at a later point in the kinetics). Thus, reducing the fragmentation rate will have less effect on the maximum aggregation rate for protein concentrations where the CFC_M is crossed significantly earlier than CFC_N, compared with those (lower concentrations) for which the two transitions are closer together. This effect becomes more apparent when the CFC_M threshold is lowered [5] because the reduced fragmentation rate k_f^* is active earlier in the kinetics. This is the reason for the difference in maximum aggregation rate for high protein concentrations in figure 6.20, compared with those in figure 6.10.

In order for these results to match the experimental data, the CFC_N must be lower than predicted by equation 6.5 such that fragmentation is arrested much closer to the CFC_M (i.e. CFC_N does not scale in the way I assumed in equation 6.5). This, therefore, presents an interesting question, since this model appears to

[5] Because of the way that I have defined them through equation 6.5, reducing the CFC_M will also lower the CFC_N.

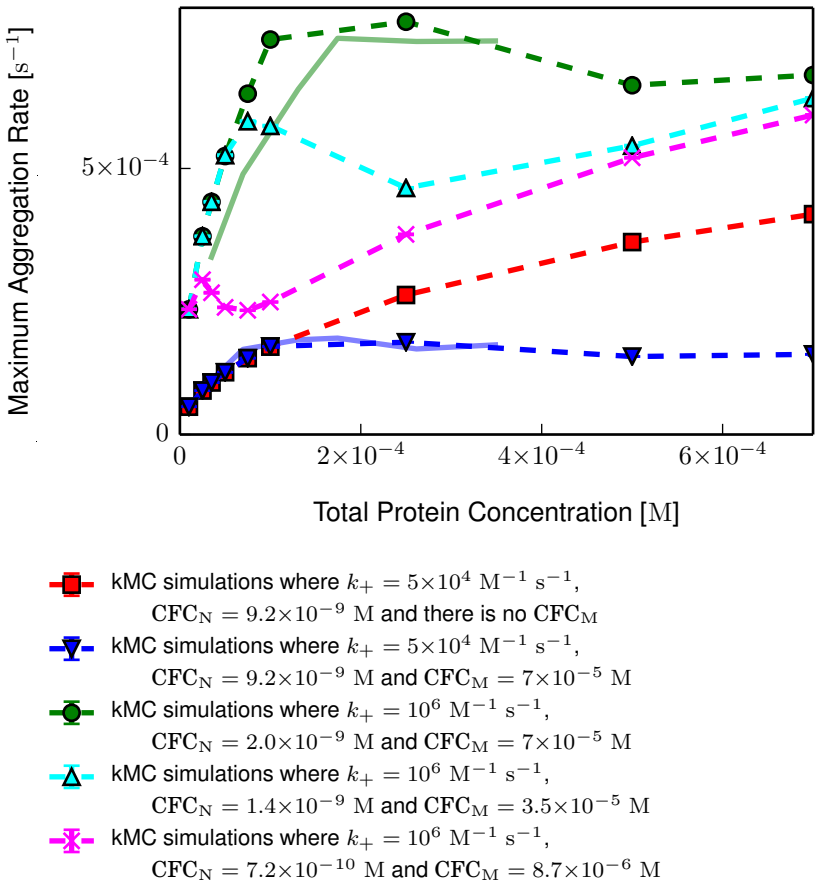


Figure 6.20 The scaling behaviour of $k_{\max}$ with m_{tot} in a fragmentation model with fragmentation slowed to $k_f^* = 3\times10^{-9}$ s^{-1} once $M(t) = \text{CFC}_\text{M}$ (if applicable), and arrested once $N(t) = \text{CFC}_\text{N} = \sqrt{2\text{CFC}_\text{M}k_f/k_+}$, for different values of CFC_M and k_+; where the common parameter values were $n_c = 2$, $N(0) = 2\times10^{-12}$ M, $k_n = 0$, $k_- = 0$, $k_f = 3\times10^{-8}$ s^{-1}, and $V = 83\times10^{-14}$ l. Each kMC simulation point is the mean of 150 replicates with error bars indicating the standard deviation (method in §2.2); the lines are to help differentiate the data sets. The bovine insulin experimental data in figure 6.2, are also shown here as **pale solid lines**, for qualitative comparison.

describe the experimental data very well in the absence of salt, but the two critical concentrations must be affected by salt in a different way, in order to reproduce all of the experimentally observed behaviour.

6.5 Discussion

It is now widely thought that short filaments (protofilaments or oligomers) are the pathogenic species in many amyloid diseases [33, 52, 62, 67, 95, 100, 112]. As such, preventing their formation is clearly important. Xue et al. [112] have also indicated that it is not only the short filaments formed during the early stages of aggregation that are toxic but also those formed through fragmentation of mature fibrils. Thus, preventing fragmentation could be effective means to reduce the number of short fibrils produced and thereby limit cell damage.

In this chapter, I have looked at a modification to the fragmentation model of filament self-assembly in which the fragmentation rate is reduced when a critical filament concentration (CFC) is reached. This model was based on the assumption that filament bundles — formed by the lateral association of filaments — have a lower probability of fragmentation than the filaments not in bundles. This I motivated by an apparent change in the kinetics and morphological properties of experimental samples, above a particular total protein concentration.

From this model it can be seen that when the rate of fragmentation is significantly reduced (perhaps to zero) the creation of new filaments is suppressed [(6)]. This means that the number of short filaments present, will be maximal at some point before, or exactly when, the CFC is reached, but can be expected to decline thereafter. Interestingly, Taylor et al. [100] have reported that the highest level of cell damage occurs during the exponential growth phase of the growth curve, with significantly lower rates of cell damage before and after this point. This could indicate that a similar transition to that which I have proposed for insulin, also suppresses fragmentation during fibril formation of amyloid-beta. If this is the case, then it should be possible to change the point in the growth curve

[(6)] This assumes that filament elongation continues to use up the monomer population, thereby preventing further primary, or secondary nucleation.

where maximal cell damage is observed, by altering the environmental conditions in such a way as to change the CFC threshold condition.

In the case of bovine insulin it appears to be possible to lower the CFC by adding increasing concentrations of salt. However, the effect that this had on the maximum aggregation rate could only be reproduced when fragmentation was actually arrested at the CFC_M; i.e. instead of merely suppressing fragmentation at the CFC_M and then arresting it at the CFC_N. This implies that the relationship between the two threshold criteria (CFC_N and CFC_M), which I assumed to be proportional to one another, may be more complicated; for example, the salt concentration perhaps changes both the threshold at which the fragmentation rate is reduced, and the value of the rate constant after the threshold is passed. This could be seen as an extension of the charge screening effect: increasing the rate of bundle formation and increasing their thickness, and thereby simultaneously lowering the threshold concentration and reducing the possibility that a bundle undergoes fragmentation.

The formation of bundles from individual amyloid filaments is a widely recognised phenomenon [61], but the actual process of lateral filament association has only been observed relatively recently [49, 66, 69] and the properties of these bundles are still largely unknown. As such, it is unclear whether all bundles can be expected to provide comparable resistance against fragmentation, or indeed how common bundle formation is for filaments made from different proteins.

The model that I presented, only implemented a coarse-grained description of the connection between the critical filament concentration, and the suppression or arrest of fragmentation. If lateral filament association were explicitly implemented in such a model it would be possible to compare the average bundle thickness with those measured experimentally [49, 60, 66]. However, this would require additional information on the properties of bundles; for example, whether bundles can join to one another, whether the addition of new filaments to a bundle depends on the bundles size, and whether filaments within bundles have a different rate of elongation to those on their own.

The rationale that salt lowers the CFC by allowing easier lateral association of filaments, was based on the assumption that the filaments had a weak Coulomb repulsion which is screened by the presence of salt. If so, then a similar effect should be produced by other factors that can influence the bundling affinity of a given protein; for example, the helical twist of the fibrils themselves [61], or effect of different solvents on a protein that contains a large number of hydrophobic amino acids. The bundling propensity for filaments of a given protein may, therefore, explain why the fibrils formed by ovalbumin under certain conditions appear more flexible than those formed by insulin (comparing figures 1.2a and b, in §1.2) or those of ovalbumin under different conditions [60, 107].

A recent study of ovalbumin fibrils by Lara et al. [60] indicates that a higher persistence length (more rigid fibrils) might be correlated with increased fibril thickness. This poses an interesting question about the internal structure of fibrils and the importance of lateral filament association. However, their findings also seem to indicate that the average length of the thicker fibrils decreases over time, while the average length of the thinner fibrils increase; i.e. clearly the opposite of the prediction that bundling prevents fragmentation. This effect is almost certainly related to filament end-joining (which I discuss in the next chapter) where the thicker filaments are not fragmenting more, but rather the thinner ones are rejoining more. Hence, there may be an interesting link between a filaments flexibility, and its tendency to fragment, end-join or laterally associate.

Chapter 7

The effects of filament end-joining and the formation of amyloid loops

7.1 Introduction

Loop-like amyloid fibrils have been observed in a number of protein systems, using a variety of imaging techniques [41, 101, 102]. As an example, figure 7.1 shows looped and non-looped fibrils of ovalbumin.

The fibrils present in samples where loops have been found, appear more flexible than in systems where loops are not reported (compare, for example, figures 1.2a and b). This is not surprising as loop formation requires that the fibrils are flexible enough that their ends are able to meet and join. Hence, the existence of loops indicates that the internal structure of some fibrils may be different, and that an alternative aggregation pathway involving end-joining is at work.

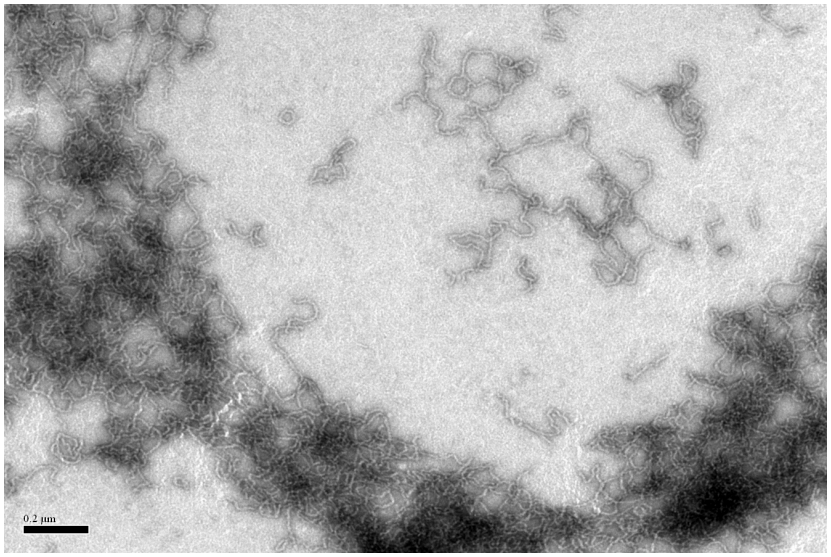


Figure 7.1 Transmission electron micrograph of amyloid fibrils and loops formed from ovalbumin. The sample was taken from experiments performed by my collaborators Ryan Morris and Jason Kalapothakis, after the fluorescence signal had plateaued. This image has a 0.2 µm scale bar.

The two processes of end-joining — where a fibril can form a loop, or grow by joining another fibril — can be described by two new reactions [(1)]:

$$f_i + f_j \xrightarrow{k_\mathrm{j}} f_{i+j} \tag{7.1}$$

$$f_i \xrightarrow{k_\mathrm{j} P_\mathrm{loop}(l)} l_i \tag{7.2}$$

where k_j is the rate of association of two filament ends and is the same in both reactions. For reasons which I will discuss shortly in §7.1.1, the loop formation reaction 7.2 also depends on the length l, of the filament, and hence the probability $P_\mathrm{loop}(l)$ that its ends find one another.

In this chapter, I will look at two models involving filament end-joining. The first, in §7.2, will consist of the reactions 7.1 and 7.2 above, as well as reaction 1.8 for elongation by monomer addition. The second model, in addition to these, will contain reactions 1.10 and 1.7 for the fragmentation of filaments (this can be regarded as reversible end-joining) and to account for the instability of filaments shorter than n_c, respectively. However, for reasons which I will now explain, the primary nucleation reaction will not be used in either model.

(1) Reaction 7.1 was previously given as reaction 1.26 in §1.3.5.

Experimental growth curves for an amyloid loop forming protein have no lag phase

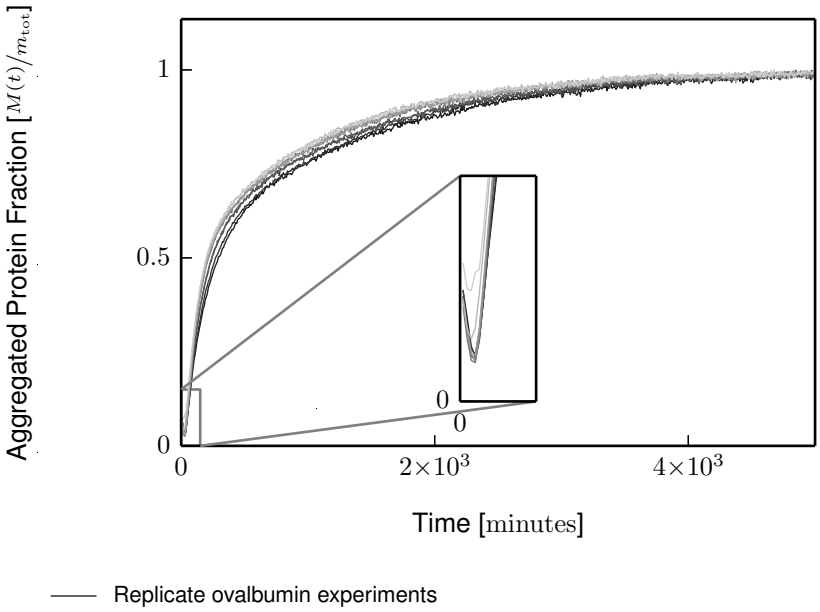


Figure 7.2 Replicate growth curves for the self-assembly of amyloid fibrils from 19 μM ovalbumin, previously shown in figure 1.12, with an inset highlighting the unexplained drop in fluorescence at the very early times (showing $0 - 150$ Minutes). These 9 replicates were obtained simultaneously, by my collaborators Ryan Morris and Jason Kalapothakis, and have been normalised by the maximum fluorescence of each curve.

Typical growth curves produced by the aggregation of ovalbumin are shown in figure 7.2. Curiously, many of the proteins which appear to form loops also produce this shape of growth curve with little or no appreciable lag phase.

The growth curves of other proteins which normally do have a lag phase, can be made to resemble those in figure 7.2 by "seeding", i.e. adding preformed fibrils to the sample at $t = 0$. This suggests that the aggregation of proteins such as ovalbumin which do not require seeding to produce growth curves without a lag phase, may involve a different nucleation pathway to the one normally considered (see §1.3.1). Since many of these proteins appear to undergo end-joining as well, it can be assumed that they have a higher affinity for aggregation in general such that these alternative nuclei form rapidly, acting like seeds. Hence, while

nucleation may be important in these kinetics, the mechanism can be reasonably approximated by seeding with a known number of fibrils.

7.1.1 The probability of filament end-joining

For a model in which two arbitrary filament ends can join, there are two scenarios that need to be considered: that the ends of two different filaments join together, and that both ends of the same filament meet and form a loop. The first of these is straightforward to model, while the second is slightly more complicated.

The probability for two filaments to end-join

There are four ways in which two filaments can combine into a single, longer filament, but I will treat these as equivalent since it is only the length of the resulting filament that will matter in my simulations [(2)]. Thus, I can write the total probability at time t for any two of the $N(t)$ filaments to meet and join, as

$$P_{\text{join}}(t) = k_{\text{j}} N(t)(N(t) - 1) \approx k_{\text{j}} N^2(t) \tag{7.3}$$

where the approximation is due to the conversion between an discrete number of filaments, and their concentration as a continuous variable.

The probability that filament forms a closed loop

In order to find the probability that a single filament forms a loop, I first need to obtain the probability that the two ends of a single filament meet. There are a number of models to describe the probability of loop formation for a polymer [42, 50, 57, 70, 114]. These treat the polymer as a chain of inflexible segments, each of length l_{p} (referred to as the persistence length), each of which may consist of a large number of monomers. Simple models treat the entire polymer as a freely jointed chain of these segments, while more sophisticated models, such as the one that I use, impose constraints on the maximum flexibility of the joints.

(2) It should be noted that I also ignore the possibility that the point where two filaments join may be weaker and hence subject to a higher probability of fragmentation than those formed by monomer addition.

While the amyloid fibrils in figure 7.1 appear more flexible than those seen for other proteins [3] they must have some rigidity. This will define the minimum fibril length $l_{\min}$ necessary for a loop to form. For this reason, I will use the worm-like chain model for polymers with finite flexibility [42, 50] to describe the mean square end-to-end distance of a polymer of length l:

$$\left\langle R^2(l)\right\rangle = 2l_{\mathrm{p}}\left(l - l_{\mathrm{p}}\left(1 - e^{-l/l_{\mathrm{p}}}\right)\right) \tag{7.4}$$

This model assumes that, under the influence of random perturbations, the polymer will move and bend but its two ends will remain on average within a spherical volume prescribed by the radius $\sqrt{\langle R^2(l)\rangle}$, known as the polymer's radius of gyration. This can be used to define the probability that the two ends are separated by a given end-to-end vector $\vec{R}(l)$ [22]:

$$P(\vec{R}(l), l) = \left(\frac{2\pi\left\langle R^2(l)\right\rangle}{3}\right)^{-3/2} \exp\left(-\frac{3\left(\left|\vec{R}(l)\right|\right)^2}{2\left\langle R^2(l)\right\rangle}\right) \tag{7.5}$$

Hence, the probability that the two ends meet can be obtained by setting $|\vec{R}(l)| = 0$ in equation 7.5 and imposing the minimum contour length $l_{\min}$:

$$P_{\mathrm{loop}}(l) = P(\vec{R}(l) = 0, l) = \begin{cases} \left(\frac{2\pi\left\langle R^2(l)\right\rangle}{3}\right)^{-3/2} & \text{if } l \geq l_{\min} \\ 0 & \text{else} \end{cases} \tag{7.6}$$

This was the probability used in reaction 7.2.

From equation 7.6 it can be seen that loop formation becomes less likely with increasing filament stiffness (persistence length) due to the increase in mean square end-to-end distance and $l_{\min}$; similarly, even for a highly flexible filament, the probability that its two ends meet will decrease as it grows, such that $P_{\mathrm{loop}}(l) \to 0$ as $l \to \infty$ [4]. From this, the total probability that any one of

[3] The insulin fibrils in figure 1.2a appear less flexible than those for ovalbumin in figure 1.2b.

[4] The ends will act as two independent random walkers in an infinite three dimensional space when $l \to \infty$.

the $N(t)$ filaments at time t forms a loop, is

$$P_{\mathrm{self-join}}(t) \sim \sum_{l=0}^{\infty} N_l(t) P_{\mathrm{loop}}(l) \tag{7.7}$$

where $N_l(t)$ is the number of filaments of length l at time t.

Using these relations, I will now consider a simple model for filament polymerisation by monomer addition and end-joining, and how this relates to the growth curves found experimentally for ovalbumin (figure 7.2).

7.2 A simple model for filament end-joining

In constructing a model comprising elongation by both end-joining and monomer addition, I first note that for a sufficiently high elongation rate, the probability of a filament forming a loop rapidly becomes negligible compared to that of two filaments joining end-to-end [(5)]. Because of this, I can neglect loop formation and simply express the depletion of filaments due to end-joining as

$$\frac{dN}{dt} = -k_{\mathrm{j}} N(t)^2$$

Using this, I construct a pair of differential equations to describe the rate of change of the fibrillar protein concentration $M(t)$ and the concentration of filaments $N(t)$, under the actions of elongation by monomer addition and end-joining, respectively:

$$\frac{dM}{dt} = 2k_{+} m(t) N(t) \tag{7.8}$$

$$\frac{dN}{dt} = -k_{\mathrm{j}} N(t)^2 \tag{7.9}$$

where k_{+} is the elongation rate, k_{j} is the rate of end-joining, and $m(t) = m_{\mathrm{tot}} - M(t)$ is the free monomer concentration at time t.

(5) This can be seen from the way that equation 7.6 scales with l: at a high rate of elongation, the fibrils will rapidly become long enough that $P_{\mathrm{loop}}(l)$ is negligible.

At this point it is worth highlighting the non-linearity that filament end-joining introduces through equation 7.9, as this will present a problem later when considered in conjunction with filament fragmentation (§7.3). Earlier in the thesis, similar non-linearity, introduced by the primary nucleation mechanism to the autocatalysis model, was neglected by assuming that the rate of primary nucleation was low and that its effect on the monomer population was small (§1.3.3). However, this cannot be done for the end-joining mechanism because the association rate k_j is unknown and, moreover, end-joining is the only means by which the number of filaments — or more accurately, the number of growth-competent ends — is depleted. For now however, this non-linearity will not cause a problem because the equations for $M(t)$ and $N(t)$ are only partially coupled; i.e. equation 7.8 depends on equation 7.9 but not *vice versa*.

As mentioned above, many of the loop forming proteins exhibit little or no lag phase in their growth curves (this was evident in figure 7.2 for ovalbumin). These are similar in appearance to high seed concentration growth curves described in §3.3.3; thus, I have ignored primary nucleation in equations 7.8 and 7.9, and will assume that the system starts with a high concentration of short filaments $N(0) = {}^{M(0)}/_{n_\mathrm{c}}$.

Using these initial conditions, I am thus able to solve equations 7.8 and 7.9 analytically and obtain the following expressions for the time evolution of the concentrations of fibrillar protein and number of filaments:

$$M(t) = m_\mathrm{tot} + (M(0) - m_\mathrm{tot})\,(1 + k_\mathrm{j}N(0)t)^{-2k_+/k_\mathrm{j}} \tag{7.10}$$

$$N(t) = \frac{N(0)}{1 + k_\mathrm{j}N(0)t} \tag{7.11}$$

7.2.1 Comparing predicted and experimental growth curves

Using the fitting technique in §2.3.1, I obtained values for the parameter combinations ${}^{2k_+}/_{k_\mathrm{j}}$, and $k_\mathrm{j}N(0)$, along with $M(0)$, by fitting equation 7.10 to the average growth curve of the ovalbumin data in figure 7.2. Although it is trivial to fit equation 7.10 for the individual parameters k_+, k_j, and $N(0)$, instead, the lack of constraints in this model means that their fitted values are sensitive to the choice of starting values (which is not the case for the combinations); indeed,

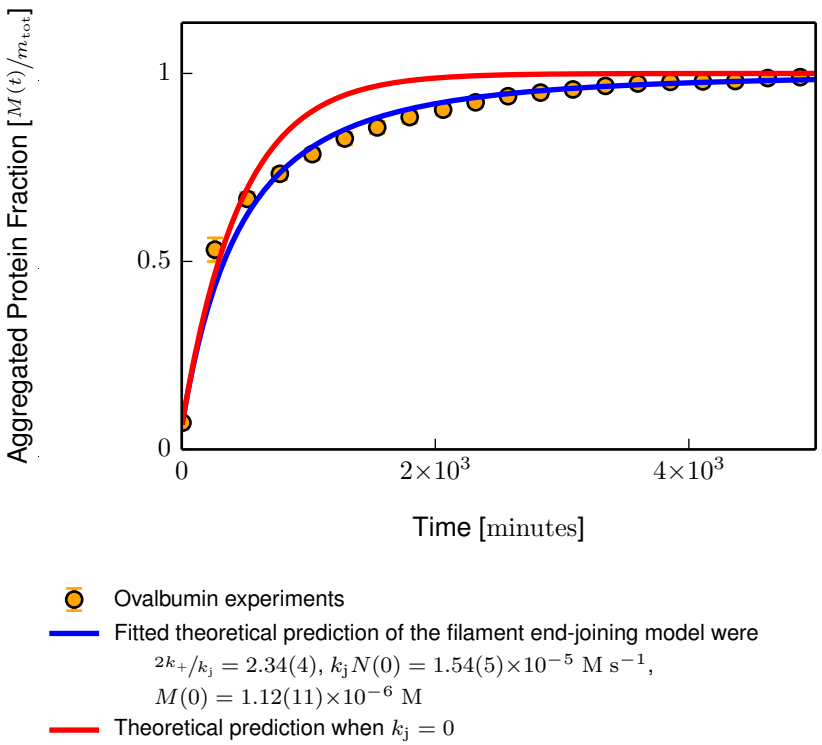


Figure 7.3 An example fit of equation 7.10 to experimental data; in this case, the mean of the replicate growth curves shown in figure 7.2 for 19 μM ovalbumin, with error bars indicating the standard deviation. Also shown is the theoretical prediction for polymerisation without end-joining.

unlike the fit for bovine insulin in §3.2, it is not possible to extract values for the individual parameters, due to the lack of additional information (for example, the mean filament length).

The best fit of equation 7.10 to the experimental data, is shown in figure 7.3, along with a growth profile for polymerisation without end-joining [(6)]. From this figure, it is clear that the end-joining model produces a better fit to the data; by comparing these growth curves, it can be seen that filament end-joining slows the rate of aggregation at intermediate times. This is due to the reduced number of growth-competent ends, which decreases the rate of elongation.

Similar slowing of the aggregation rate can also be seen in the experimental growth curves, indicating that filament end-joining is indeed involved. However,

(6) An expression for polymerisation without end-joining can be found by solving equation 7.8 with time independent N; here, I have plotted this solution, making use of the fitted parameter values for $2k_+N(0)$.

it appears that this may be slightly more pronounced in the experimental data than predicted by the model; i.e. the rate of aggregation in the experimental data at very early times, appears slightly higher and then turns over slightly more rapidly.

The shallower curvature of the intermediate section of the experimental growth curve suggests that the number of ends changes less rapidly here, than at early times. This behaviour is indicative of a non-zero asymptotic limit to the minimum number of growth-competent ends present; i.e. if left unchecked, filament end-joining would result in zero ends — all filaments having irreversibly formed into closed loops. However, here it appears that end-joining cannot reduce the number of ends beyond an "equilibrium" value. One method by which this can be achieved is the introduction of fragmentation into the model, making filament end-joining reversible.

7.3 A mean-field model for reversible filament end-joining

In the previous section, I considered a model containing only monomer addition and filament end-joining, which resulted in an easily separable pair of differential equations for the concentrations of fibrillar protein $M(t)$ and number of filaments $N(t)$. However, combining end-joining and fragmentation into a single model, renders the model difficult to solve analytically; the equivalent pair of equations with the inclusion of fragmentation (reaction 1.10) are no longer separable:

$$\frac{dM}{dt} = 2k_{+}m(t)N(t) - k_{\mathrm{f}}n_{\mathrm{c}}(n_{\mathrm{c}} - 1)N(t) \tag{7.12}$$

$$\frac{dN}{dt} = -k_{\mathrm{j}}N(t)^2 + k_{\mathrm{f}}\left(M(t) - (2n_{\mathrm{c}} - 1)\,N(t)\right) \tag{7.13}$$

where k_+ is the elongation rate, and $m(t) = m_{\mathrm{tot}} - M(t)$ is the free monomer concentration. The first term in each equation also appeared in the simpler model [7], and the second term in each equation (filament fragmentation with rate constant k_f) is the same as appeared in the fragmentation model (equations 1.12 and 1.13, in §1.3.3); again, I have ignored primary nucleation.

Starting from equations 7.12 and 7.13, I will now describe the method by which closed-form analytic expressions for $M(t)$ and $N(t)$ were obtained [8] for a model describing monomer addition and reversible filament end-joining.

7.3.1 Obtaining a closed-form analytic solution

Careful examination of equations 7.12 and 7.13 reveals two physically meaningful fixed points: one at

$$\begin{pmatrix} M^* \\ N^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{7.14}$$

which is unstable and acts as a source; and one at

$$\begin{pmatrix} M^* \\ N^* \end{pmatrix} = \begin{pmatrix} m_{\mathrm{tot}} - \frac{k_\mathrm{f} n_\mathrm{c}(n_\mathrm{c}-1)}{2k_+} \\ \frac{-k_\mathrm{f}(2n_\mathrm{c}-1)+\sqrt{k_\mathrm{f}^2(2n_\mathrm{c}-1)^2+4k_\mathrm{j}k_\mathrm{f}M^*}}{2k_\mathrm{j}} \end{pmatrix} \tag{7.15}$$

which is stable and describes the steady state in the long-time limit [9].

When the rate of fragmentation is significantly lower than the rates of either elongation or filament end-joining (i.e. $k_\mathrm{f} \ll k_+, k_\mathrm{j}$), the stable point (equation 7.15)

[7] These terms described the rate of change of the fibrillar protein concentration $M(t)$ (from equation 7.8) and the concentration of filaments $N(t)$ (from equation 7.9), under the actions of elongation by monomer addition and end-joining, respectively.

[8] This process is described in more detail in §A3, and was identified by my collaborator Juraj Szavitz-Nossan.

[9] Another fixed point exists for $M^* < 0$, which will not be considered for obvious reasons.

can be approximated to

$$\begin{pmatrix} M^* \\ N^* \end{pmatrix} \approx \begin{pmatrix} m_{\text{tot}} \\ \sqrt{\frac{k_{\text{f}} m_{\text{tot}}}{k_{\text{j}}}} \end{pmatrix} \tag{7.16}$$

From this, it can be seen that the steady-state value M^* is simply, as expected, the point where all protein has aggregated, and the steady-state value N^* is determined by a dynamic equilibrium between the creation of filaments through fragmentation and their loss through filament end-joining.

Equations 7.12 and 7.13 can now be simplified by rewriting them in terms of the dimensionless variables $m(\tau) = {}^{M(t)}/_{M^*}$ and $n(\tau) = {}^{N(t)}/_{N^*}$, making use of the simplified steady-state fixed point:

$$\frac{dm(\tau)}{d\tau} = (1 - m(\tau))n(\tau) \tag{7.17}$$

$$\frac{dn(\tau)}{d\tau} = -\, r\, n(\tau)^2 + r m(\tau) \tag{7.18}$$

where terms $\mathcal{O}({}^{k_{\text{f}}}/_{k_+})$ and $\mathcal{O}({}^{k_{\text{f}}}/_{k_{\text{j}}})$ have been neglected, $r = {}^{k_{\text{j}}}/_{2k_+}$ is the dimensionless ratio of the rates of filament end-joining and elongation, and time has been rescaled to the dimensionless parameter $\tau = 2k_+ N^* t$. Interestingly, ${}^{t}/_{\tau}$ is related to the characteristic time scale κ^{-1} via $\sqrt{r}$:

$$\frac{t}{\tau} = \frac{1}{2k_+}\sqrt{\frac{k_{\text{j}}}{k_{\text{f}} m_{\text{tot}}}} = \frac{\sqrt{r}}{\kappa} \tag{7.19}$$

where I have used equation 7.16 to approximate N^*.

An approximate solution to equation 7.18 can be found by first solving for $m(\tau)$ in the long-time limit when $n \to 1$. Here, $m(\tau)$ takes the form of an exponential approach from $m_0 = {}^{M(0)}/_{M^*}$, to the steady state $m(\infty) = 1$. Using this approximation to $m(\tau)$, the time evolution of $n(\tau)$ can thus be found:

$$n(\tau) = \frac{B(2r\sqrt{1 - m_0}e^{-\tau/2})}{2r} \tag{7.20}$$

where the parameter $B(z)$ is defined in terms of the Bessel functions $J_\nu(z)$ and $Y_\nu(z)$, and the constant b set by the initial conditions m_0 and n_0. From

this approximation of $n(\tau)$, a solution to equation 7.17 for $m(\tau)$, can now be found by evaluating

$$m(\tau) = 1 - (1 - m_0)e^{-\int_0^\tau n(\tau)} \tag{7.21}$$

using equation 7.20 to obtain

$$m(\tau) = 1-(1-m_0)\left(\frac{cJ_{2r}(2r\sqrt{1-m_0}) + Y_{2r}(2r\sqrt{1-m_0})}{cJ_{2r}(2r\sqrt{1-m_0}e^{-\tau/2}) + Y_{2r}(2r\sqrt{1-m_0}e^{-\tau/2})}\right)^{1/r} \tag{7.22}$$

Equation 7.22 thus describes the evolution of the fibrillar protein concentration in terms of dimensionless quantities $m_0 = {}^{M(0)}/_{M^*}$, $n_0 = {}^{N(0)}/_{M^*}$, $r = {}^{k_\mathrm{j}}/_{2k_+}$ and $\tau = 2k_+N^*t$, where $M^* \approx m_\mathrm{tot}$ and $N^* \approx \sqrt{\frac{k_\mathrm{f}m_\mathrm{tot}}{k_\mathrm{j}}}$. I will now explore the predictions of this solution and how they compare with kinetic Monte Carlo simulations of a model consisting of monomer addition, filament breakage and reassociation, and where filaments shorter than n_c disintegrate (reactions 1.8,1.10, 7.1, and 1.7, respectively).

7.3.2 Comparing the mean-field prediction to kinetic Monte Carlo simulation

The assumptions made while obtaining equation 7.22 were that fragmentation was rare compared to elongation and filament end-joining (i.e. $k_\mathrm{f} \ll k_+, k_\mathrm{j}$), and that $n(\tau)$ could be adequately described by approximating $m(\tau)$ as an exponential approach to the steady state; i.e. solving equation 7.18 for $m(\tau)$ in the limit $n \to 1$, then using this to solve equation 7.18. The assumption that $m(\tau)$ can be approximated as exponential approach means that this solution is only valid when the initial rate of aggregation is the maximum rate of aggregation; indicating that the seed concentration $N(0)$ must be high, and in fact $N(0) \geq N^*$. This constrains the maximal value for the rate of fragmentation, through the definition of $N^* \propto \sqrt{k_\mathrm{f}}$ in equation 7.16, thus ensuring that the first of the two above listed approximations is also fulfilled.

For completeness, kinetic Monte Carlo simulation results will be presented for both $N(0) < N^*$ and $N(0) \geq N^*$, but for the reasons given above, equation 7.22 will only be considered in conjunction with the latter case.

The effects of a low seed concentration: $N(0) < N^*$

When few seeds are present — perhaps because of a low rate of nucleation — fragmentation will dominate the growth curve in the same way as it did in the autocatalytic model (§3).

At early times there is a low probability of filament end-joining, due to the low population of filaments. Fragmentation also has a low probability at this time, until a sufficient amount of material has been accreted. Once this happens and new filaments are produced through fragmentation, the probability that they recombine also increases. Consequently, the early-time kinetics are expected to be as they were in the autocatalysis model, but the steady state equilibrium between breakage and rejoining means that fewer growth-competent ends will be present at later times. Hence, the growth curves are expected to appear different to those produced by the fragmentation model without end-joining in §3.3.

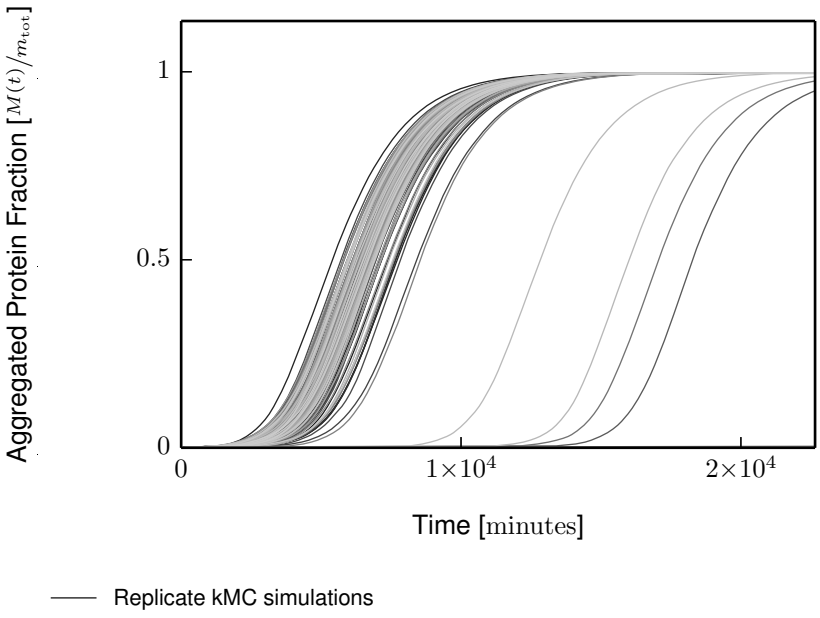


Figure 7.4 Growth curves produced by kMC simulations of a reversible filament end-joining model for amyloid fibril self-assembly, where the parameter values were $m_{\rm tot} = 10\ \mu$M, $n_{\rm c} = 2$, $N(0) = 1 \ll N^*$, $k_{\rm n} = 0$, $k_+ = 5{\times}10^3\ {\rm M}^{-1}\ {\rm s}^{-1}$, $k_- = 0$, $k_{\rm j} = 5{\times}10^4\ {\rm M}^{-1}\ {\rm s}^{-1}$, $k_{\rm f} = 5{\times}10^{-9}\ {\rm s}^{-1}$, $V = 83{\times}10^{-14}$ l, and for a monomer size of 26 Å, $l_{\rm p} = 36$ nm and $l_{\rm min} = 205$ nm. Of the 150 replicate simulations, 3 posses lag times exceeding 10^6 minutes, and therefore have not been shown.

The growth curves produced by kinetic Monte Carlo simulations of this model with a low seed concentration are shown in figure 7.4. These are indeed sigmoidal and somewhat reminiscent of the growth curves produced by the heterogeneous nucleation model (see figure 3.5 in §3.3.2). This similarity is due the restriction on the number of filaments present [(10)] having much the same affect as monomer depletion did on the heterogeneous nucleation of new filaments.

When $k_+ < k_j$, the steady state number of filaments N^* at the point where the rate of aggregation is maximal, will be lower than the number of fibrils present in the fragmentation model. As a result, the maximum aggregation rate, which previously scaled with κ in the autocatalysis model (without end-joining), will be limited by the characteristic rate $\kappa/\sqrt{r}$, according the equilibrium number of filaments.

In spite of not including a nucleation step, there is a surprisingly large variability in lag time between the growth curves in figure 7.4. This variability is due to the high rate of filament end-joining which causes increased fluctuations in the small number of growth-competent filaments at early times.

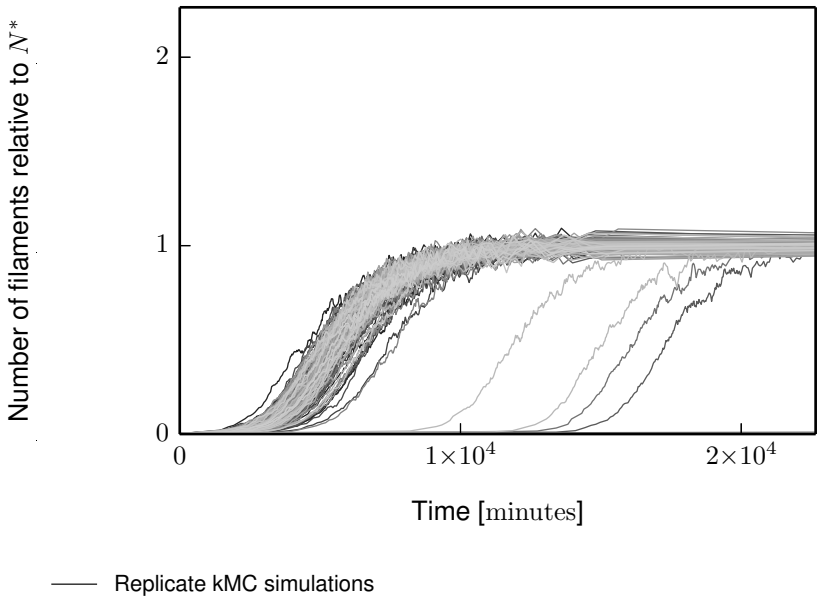


Figure 7.5 The time evolution of the number of filaments in the kMC simulation growth curves shown in figure 7.4. These have been normalised by the predicted equilibrium number of filament, N^*.

[(10)] Those filaments formed through fragmentation, recombine, limiting the number of filaments present, to N^* in the long-time limit.

The time evolution of the number of filaments according to kinetic Monte Carlo simulations are shown in figure 7.5. As expected, the number of filaments tends to N^* in the long-time limit.

The effects of a high seed concentration: $N(0) \geq N^*$

When a large number of short filaments are present at early times the rate of fragmentation will be negligible compared to the rate of filament end-joining. Initially, this will result in a rapid decrease in the number of growth-competent ends as it did in the simpler model without fragmentation. However, once the total fibrillar protein concentration is sufficiently high, and the number of filaments has been sufficiently reduced, the rate of fragmentation will become comparable with that of rejoining, thus establishing an equilibrium in the number of filaments.

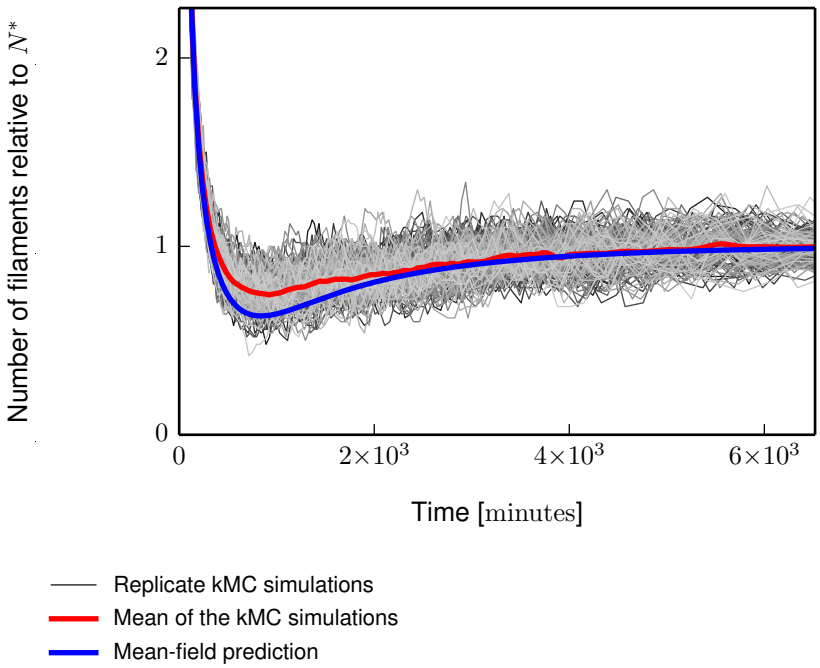


Figure 7.6 The time evolution of the number of filaments in the kMC simulation growth curves shown in figure 7.7, along with the mean-field prediction given in equation 7.20. These have been normalised by the predicted equilibrium number of filaments, N^*. The average of the simulations was obtained from the 150 replicates shown, by interpolating each curve into continuous time and then computing the mean $\langle N(t) \rangle$, at discrete time intervals.

From figure 7.6, it can be seen that fragmentation first becomes comparable with rejoining, relatively early in the kinetics, and that the equilibrium established is unstable: this is indicated by the existence of a minimum in the curve. Once the equilibrium is established, the rate of fragmentation increases as the concentration of fibrillar protein increases, and the number of filaments tends to the long-time steady state value N^*.

This behaviour is reproduced fairly well by the mean-field expression (equation 7.20), although the minimum in $N(t)$ (in figure 7.6) is lower than the average for the simulations. This discrepancy arises from neglecting terms $\mathcal{O}(k_\mathrm{f})$ in the derivation of equation 7.20, which become important only in establishing the equilibrium by damping the rapid decrease in the number of filaments. Because of this, when k_f becomes vanishingly small, agreement between equation 7.20 and the simulation results will improve [(11)].

Figure 7.7 shows the growth curves (i.e. $M(t)$) corresponding to the data shown in figure 7.6. From these it can be seen that the minimum in the number of filaments (in figure 7.6) is responsible for an abrupt change in aggregation rate just after the early-time rapid growth (in figure 7.7). Here, the gradual increase in the number of filaments that can be seen in figure 7.6 balances the depletion of the monomer population, thereby reducing the curvature of the growth curve. These simulated growth curves are similar in appearance to the experimental growth curves of ovalbumin displayed in figure 7.2.

Because equation 7.20 underestimates the number of filaments present during this quasi-linear portion of the growth curve, equation 7.22 underestimates the total monomer content of these filaments during the same time period. This causes the discrepancy between the mean-field prediction and the average of the simulation results at around 10^3 minutes in both figures 7.6 and 7.7.

The significantly lower variability between the replicate curves for $M(t)$, compared to those for $N(t)$, is caused by two factors. Firstly, the threshold number of filaments is much smaller than the total number of monomers, so a change in the number of filaments is more apparent than a change in their total monomer content. The other factor is the negative feedback between fragmentation and rejoining, where small perturbations about the equilibrium number of filaments at

[(11)] In the limit where $k_\mathrm{f} \to 0$, this model will recover the prediction of the simple model in §7.2.1.

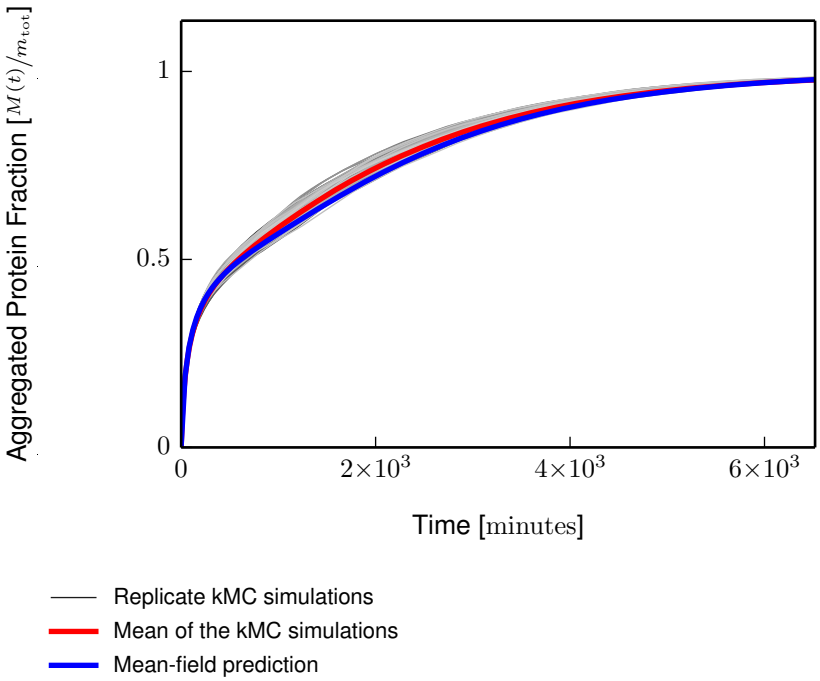


Figure 7.7 Growth curves produced by kMC simulations of the reversible filament end-joining model for amyloid fibril self-assembly, along with the mean-field prediction given in equation 7.22, where the parameter values were $m_{\rm tot} = 10$ μM, $n_{\rm c} = 2$, $N(0) = 750 \gg N^*$, $k_{\rm n} = 0$, $k_+ = 5{\times}10^3$ M^{-1} s^{-1}, $k_- = 0$, $k_{\rm j} = 5{\times}10^4$ M^{-1} s^{-1}, $k_{\rm f} = 5{\times}10^{-9}$ s^{-1}, $V = 83{\times}10^{-15}$ l, and for a monomer size of 26 Å, $l_{\rm p} = 36$ nm and $l_{\rm min} = 205$ nm. The average of the simulations was obtained from the 150 replicates shown, using the method in §2.2.

a particular time by one of these two mechanisms are corrected by the other. If the correction overshoots the equilibrium value, a feedback condition is established and the number of filaments begins to oscillate about the equilibrium value.

7.3.3 Loop formation and its effect on the mean filament length

Loop formation was included in the simulations using reaction 7.2, but was not taken into account in the analytic prediction. As expected, this appears to have had a negligible impact on the kinetics, which depend on the equilibrium in the number of filament ends and their combined monomer content, and

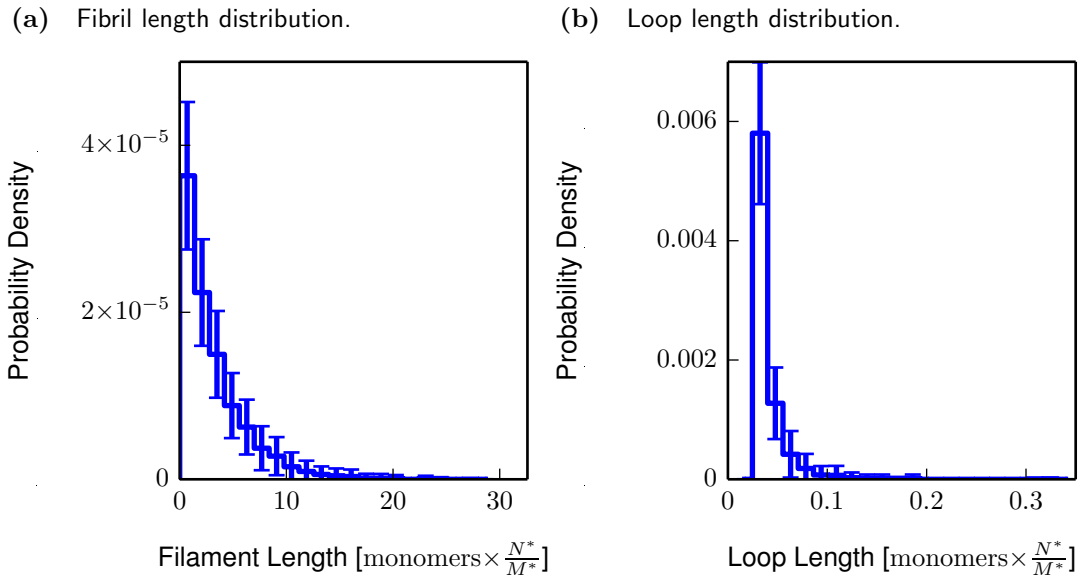


Figure 7.8 The average distribution of **(a)** filament and **(b)** loop lengths at the end-point of the replicate kMC simulation growth curves shown in figure 7.7. Each distribution represents the mean of 150 replicate simulations with error bars indicating the standard deviation (determined using the method in §2.2), and has been normalised by the predicted mean filament length, equation 7.23. The minimum filament length required to form a loop causes an offset from the origin in the loop length distribution.

not significantly on how the ends and monomer content are distributed among individual filaments.

The extent of the difference between the simulation and analytic prediction can be understood by considering the total probability for fragmentation, which is proportional to the total number of monomer-monomer bonds: a filament has one fewer of these bonds than it has monomers, whereas for a loop the number of monomers and bonds is equal. Consequently the probability of fragmentation in the kinetic Monte Carlo simulations will be higher than is predicted by the analytic solution, by a factor proportional to the number of loops in the system. This number is always smaller, and usually much smaller, than the total number of monomers in fibrils, and hence the effect is negligible.

The mean filament length at time t is given by $M(t)/N(t)$ (equation 1.25 in §1.3.4). For the reversible end-joining model, this can trivially be obtained in the long-time limit (i.e. M^*/N^*) by using the approximate values for this steady state

(equation 7.16):

$$\lim_{t\to\infty} \overline{L}(t) = \sqrt{\frac{m_{\mathrm{tot}} k_{\mathrm{j}}}{k_{\mathrm{f}}}} \tag{7.23}$$

However, this does not account for loops, which contribute to $M(t)$ but not to $N(t)$ [(12)], thereby reducing the mean filament length. In order to obtain the true mean filament length, the concentration of protein in loops at equilibrium (M^*_{loops}) would also be needed:

$$\lim_{t\to\infty} \overline{L}(t) = (m_{\mathrm{tot}} - M^*_{\mathrm{loops}}) \sqrt{\frac{k_{\mathrm{j}}}{k_{\mathrm{f}} m_{\mathrm{tot}}}}$$

As can be seen from figure 7.8, this correction is small because the average loop length is much smaller than the average filament length. This is why the mean filament length in figure 7.8a appears indistinguishable from the prediction, despite having neglected loop formation in the analytic solution.

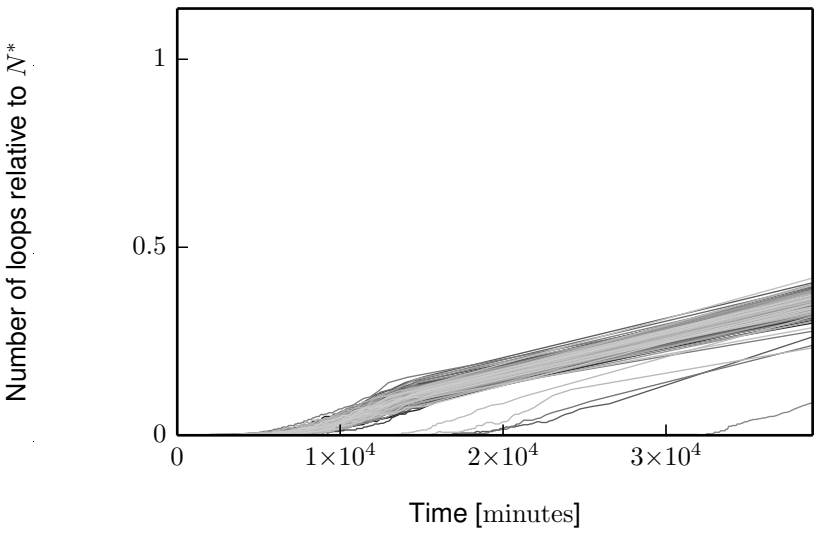
While the mean loop length is small by comparison to the mean filament length, the number of loops at the end-point is comparable with the number of filaments N^*. In the simulations, loop formation is reversible so a stable equilibrium is expected to become established between the numbers of loops and filaments.

The number of loops present in the kinetic Monte Carlo simulations are shown in figures 7.9a and b, for $N(0) < N^*$ and $N(0) \geq N^*$, respectively. The time range shown in the plots was chosen to ensure that the equilibrium values of M^* and N^* had been reached, but it is apparent from both of these figures that the number of loops has not equilibrated by this time.

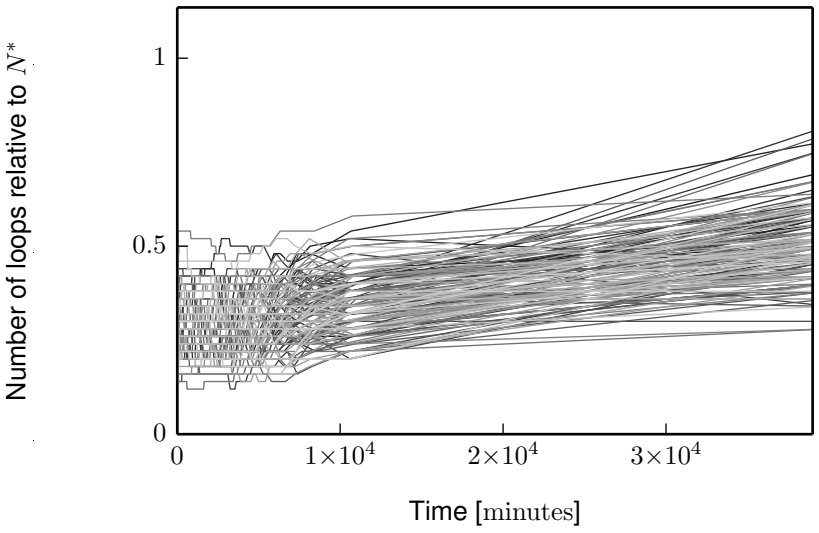
When $N(0) \geq N^*$, the large number of filaments initially available results in a higher rate of loop formation at early times. This causes the appearance of a large number of loops early on in the kinetics (see figure 7.9b). This number remains fairly constant throughout aggregation and only starts increasing further once aggregation is close to completion; i.e. when fragmentation begins to cause the disintegration of filaments into shorter pieces which have a higher probability $P_{\mathrm{loop}}(l)$ of forming a loop. By contrast, when $N(0) < N^*$, loops only

(12) Loops contribute to the total aggregated protein concentration, but not to the number of growth-competent filaments.

(a) When the initial number of filaments is below the equilibrium value (i.e. for the growth curves in figure 7.4).



(b) When the initial number of filaments is above the equilibrium value (i.e. for the growth curves in figure 7.7).



—— Replicate kMC simulations

Figure 7.9 The time evolution of the number of loops in the kMC simulation growth curves shown in figures 7.4 and 7.7; when **(a)** $N(0) \ll N^*$, or **(b)** $N(0) \gg N^*$. These have been normalised by the predicted equilibrium number of filaments, N^*.

emerge later in the kinetics, again when the number of filaments has started to noticeably increase due to fragmentation.

7.4 Predictions of a reversible filament end-joining model applied to ovalbumin

As I demonstrated in §7.3.2, a model comprising filament elongation, breakages and rejoining, and started with a high seed concentration, produces growth curves similar in appearance to those of ovalbumin in figure 7.2; these display rapid aggregation initially, but abruptly slowing to quasi-linear growth that is not found in the absence of filament breakage (see §7.2). Using this model in a similar way to the approach presented in §3.2, I have obtained fitted parameter values from the experimental data of ovalbumin, using equation 7.22 for $m(\tau)$ in §7.3.1.

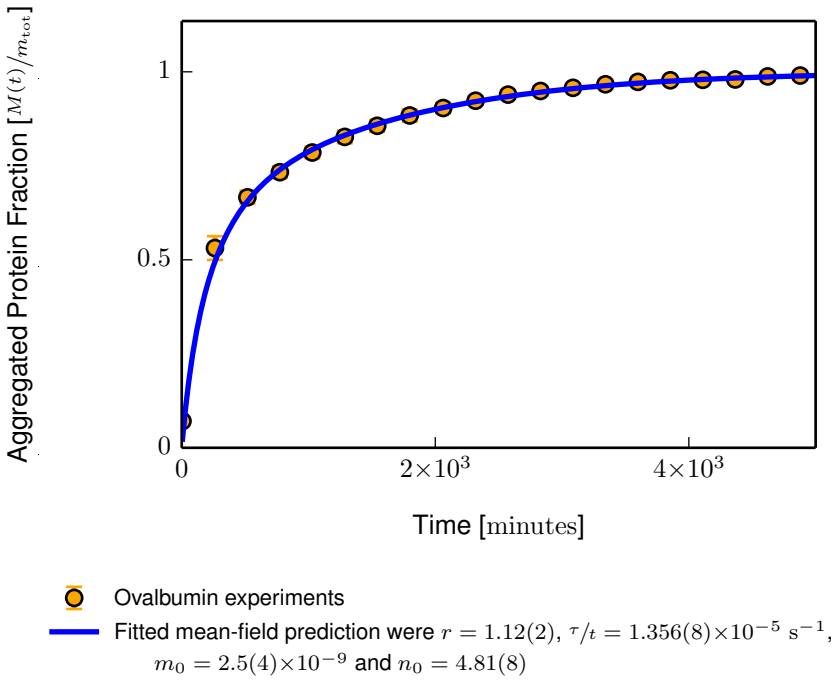


Figure 7.10 An example fit of equation 7.22 to experimental data; in this case, the mean of the replicate growth curves shown in figure 7.2 for 19 μM ovalbumin, with error bars indicating the standard deviation.

As shown in figure 7.10, the quality of the fit for the model with fragmentation and end-joining is much better than that of the simpler model without fragmentation (figure 7.3), and, while this does not give direct access to the rate constants, the fit parameters are far more robust to the choice of starting values. Values of these fitted parameters similar to those given in figure 7.10, are were also found for other concentrations of ovalbumin.

The value of $r \approx 1$ is particularly interesting as it indicates that the rates of elongation and filament end-joining are comparable; indeed, it suggests that the rate of end-joining is roughly double that of elongation. This may imply that a free monomer must undergo some sort of conformational rearrangement before, or upon docking, whereas filament ends already posses the necessary β-sheet conformation and can thus join more rapidly.

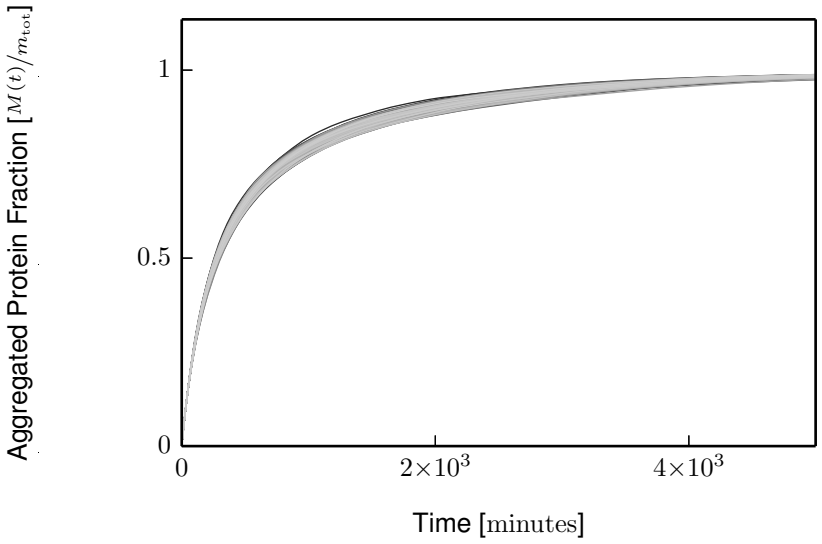
Unlike the analysis of bovine insulin in §3.2, the mean filament length is not experimentally known for ovalbumin, due to the high degree of filament and loop entanglement that is apparent in figure 7.1. Because of this, the rate constants and initial conditions cannot be extracted from the fitted parameter values without first defining one of the rates. Therefore, so as to perform kinetic Monte Carlo simulations of these parameters, I will choose a slower rate of elongation than that found for bovine insulin (see §3.2) which appeared to have a much higher overall rate of aggregation.

7.4.1 Stochastic simulations of ovalbumin aggregation

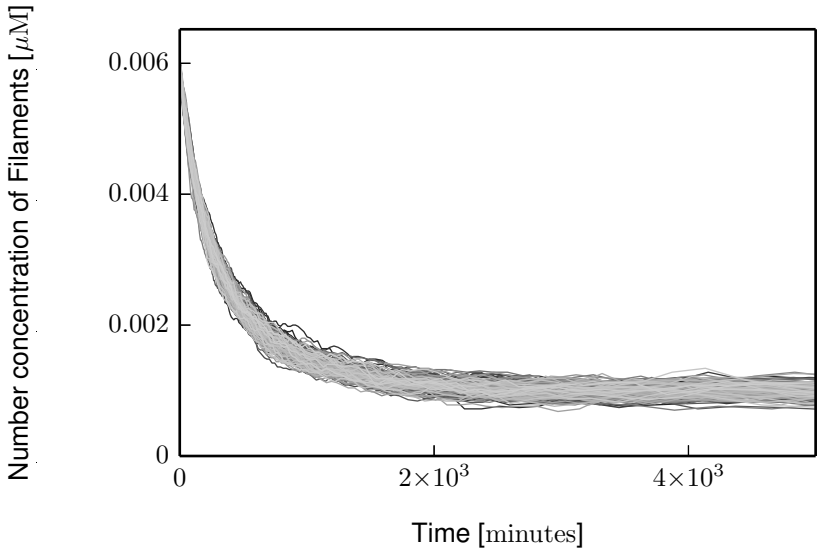
Given a value for the elongation rate k_+, the combination of parameters $2k_+N^*$ and $\frac{N(0)}{N^*}$, allow the initial concentration of fibrils $N(0)$ to be calculated. Since $M^* \approx m_{\text{tot}}$ (from 7.16), the parameter $\frac{M(0)}{M^*}$ should also provide information on the initial concentration of filaments; however, the normalised fluorescence signal at $t = 0$ is much higher than $2.5{\times}10^{-9}$ (i.e. the fitted value of $\frac{M(0)}{M^*}$ from figure 7.10), and exhibits unexplained behaviour at very early times as illustrated by the inset in figure 7.2. For this reason, and because the kinetics are more sensitive to $N(0)$ than $M(0)$, I have used $M(0) = n_{\text{c}}N(0)$ to define the initial conditions from $2k_+N^*$ and $\frac{N(0)}{N^*}$.

This gives the rate constant and initial condition values $k_+ = 5{\times}10^3$ M^{-1} s^{-1}, $k_{\text{j}} = 1{\times}10^4$ M^{-1} s^{-1}, $k_{\text{f}} = 1{\times}10^{-9}$ s^{-1}, $N(0) = 6{\times}10^{-3}$ μM. These are physically realistic, but are unlikely to be the exact values of the relevant parameters as I have had to assume one of the values.

(a) Replicate growth curves.



(b) The concentration of filaments present in the growth curves above.



—— Replicate kMC simulations

Figure 7.11 The growth curves **(a)**, and the corresponding time evolution of the concentration of filaments **(b)**, produced by 150 replicate kMC simulations of the reversible filament end-joining model for amyloid fibril self-assembly, where the parameter values were $m_{\rm tot} = 10$ μM, $n_{\rm c} = 2$, $N(0) = 300 \gg N^*$, $k_{\rm n} = 0$, $k_+ = 5{\times}10^3$ M^{-1} s^{-1}, $k_- = 0$, $k_{\rm j} = 1{\times}10^5$ M^{-1} s^{-1}, $k_{\rm f} = 1{\times}10^{-9}$ s^{-1}, $V = 83{\times}10^{-15}$ l, and for a monomer size of 26 Å, $l_{\rm p} = 36$ nm and $l_{\rm min} = 205$ nm.

Growth curves

Examples of kinetic Monte Carlo simulations performed using the rates and initial conditions determined above, are shown in figures 7.11a and b, for the growth curves and evolution of the number of filaments, respectively. These are reasonably independent of the choice of elongation rate since all other rates and the initial conditions rescale accordingly: only the scale of the time axis is affected.

These growth curves have a similar degree of variability to those in figure 7.2, found experimentally. However, these simulations are for a volume many orders of magnitude smaller than was used for the experiments. Thus, following on from my discussed in §4, it would appear that the experimental kinetics have substantially higher variability than can be accounted for simply by a difference in the number of filaments present.

Filament and loop length distributions

As mentioned previously, the length distribution is hard to determine experimentally due to the high degree of filament and loop entanglement. Recently a length distribution for ovalbumin was published by Lara et al. [60] but did not provide a corresponding distribution of loop lengths. Consequently, in order to verify that the loop length distributions produced by the kinetic Monte Carlo simulations are realistic I will compare them qualitatively to the distributions of loop lengths for another amyloid loop forming protein, Apolipoprotein C-II (apoC-II), which were reported by Hatters et al. [42] and Yang et al. [114].

The length distributions for both loops and filaments found at the end-point of kinetic Monte Carlo simulations of a reversible end-joining model, are shown together in figure 7.12 with lengths given in terms of the modified mean filament length (equation 7.23).

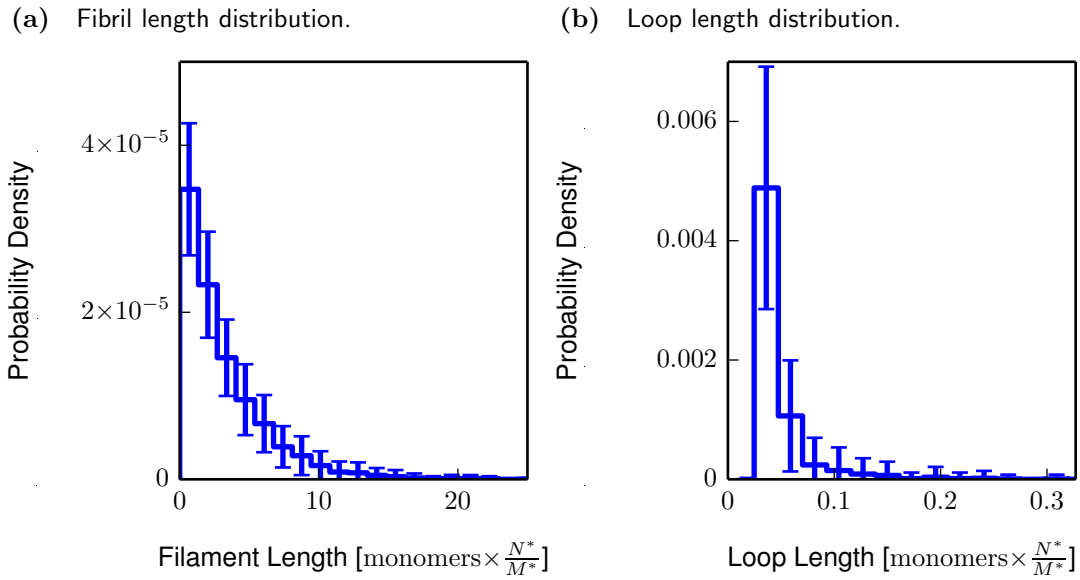


Figure 7.12 The average distribution of **(a)** filament and **(b)** loop lengths at the end-point of the replicate kMC simulation growth curves shown in figure 7.11. Each distribution represents the mean of 150 replicate simulations with error bars indicating the standard deviation (determined using the method in §2.2), and has been normalised by the predicted mean filament length, equation 7.23. The minimum filament length required to form a loop causes an offset from the origin in the loop length distribution.

The filament length distribution (figure 7.12a) is similar in appearance to the one produced (in the long-time limit) by the fragmentation model without end-joining (see §3.3.6). This is because longer filaments and loops have a higher probability of breaking (thus forming a larger population of short filaments). However, this distribution does not collapse in the way seen for that model, due to the equilibrium between filaments breaking and rejoining (since there are more short filaments, they have a high probability of rejoining). This monotonically decaying distribution of filament lengths is almost identical to that found by Lara et al. [60] for worm-like fibrils of ovalbumin.

In contrast, the loop length distribution (figure 7.12b) is mostly affected by the high population of short filaments from which to form loops and the shape of the probability function $P_{\text{loop}}(l)$ (equation 7.6). This length distribution is qualitatively very similar to that reported by Hatters et al. [42] (power law), but very different from that reported by Yang et al. [114] (a skew normal distribution). The distribution that I obtain is a steeply decaying function

with a sharp peak at the minimum loop contour length, strongly resembling the probability function for loop closure $P_{\text{loop}}(l)$ (equation 7.6). This function, however, decays less steeply than does the loop length distribution. This steeper distribution is the combined result of the decreasing probability of loop formation with length [(13)], and the underlying filament length distribution [(14)] in figure 7.12a: the high population of short filaments present at the end-point will readily form loops, whereas the sparse population of longer filaments will not.

These results suggest that the underlying filament length distributions of the studies performed by both Hatters et al. [42] and Yang et al. [114] may be qualitatively different: the former being a decaying distribution similar to figure 7.12a, while the latter may be a skew normal distribution reminiscent of the fragmentation model with fragmentation arrested at a critical fibril concentration (see §6). This could be supported by the fact that Lara et al. [60] observe an exponential filament length distribution for worm-like ovalbumin fibrils, but a peaked distribution for straighter ovalbumin fibrils (these also display higher fragmentation). Hence, similar differences in the structure of apoC-II fibrils could lead to similar filament length distributions.

Thus, it should be possible to determine an appropriate expression for the probability of loop formation (i.e. equation 7.6) by comparing experimentally measured length distributions of both filaments and loops found in the same sample.

7.5 Discussion

While the models which I have presented in this chapter are not the first attempt at understanding the significance of loop formation in the kinetics of amyloid fibril formation [5, 41, 101, 114], they do provide closed-form analytic descriptions of the time evolution of the concentrations of fibrillar protein and number of filaments.

(13) The higher probability of that short filaments form loops, coupled with the minimum contour length l_{min} defines the prominent peak in figure 7.12b.

(14) The probability for loop formation scales with the available number of filaments of a particular length.

The simple model, in which end-joining was not reversible, provided an adequate description of the growth curves experientially observed for ovalbumin. However, in these experimental data I noted a subtle departure from the prediction that I argued could only be accounted for by an equilibrium state in the number of growth-competent ends, with a non-zero value. By incorporating fragmentation into the model, this equilibrium condition could be satisfied.

The resulting model for reversible end-joining predicts that, when the initial number of filaments is much higher than the equilibrium number, the growth curves will have a high initial aggregation rate that rapidly decreases, leading to quasi-linear growth. Growth curves similar in appearance to these — where the maximum rate of aggregation occurs at the start — have been produced experimentally by a number of other proteins, including amyloid beta (Aβ) [4, 29, 30, 79, 98]. While these may only indicate that a high seed concentration is present, there are a few for which loop formation has also been observed, including alpha-synuclein (α-syn) and Apolipoprotein C-II (apoC-II) [41, 51, 102]. As such this model is expected to be reasonably widely applicable to the self-assembly of a number of different fibril forming proteins.

Unfortunately, the analytic solution obtained for this model can only be used in cases where the initial seed concentration of filament ends is higher than the equilibrium steady state value. However, the model does not rule out the possibility that sigmoidal growth curves exist in conjunction with end-joining, and using kinetic Monte Carlo simulations I demonstrated that growth curves with sub-critical concentration of seeds may appear very similar to those predicted for heterogeneous nucleation. This, in conjunction with the previous remark, suggests that there could be many other proteins for which fibril end-joining is an important process, but that these have yet to be identified.

While a number of proteins have been seen to form amyloid loops, these may be unusual for being both flexible and able to end-join. As I speculated in the previous chapter, fibril flexibility may only indicate that the filaments are not able to associate laterally into bundles, which would imply that flexible filaments might be found in the absence of loops (i.e. flexible filaments that are unable to end-join). Unfortunately, this is hard to test because failure to find loops does not mean that they are not present, particularly since flexible filaments are often found more entangled. That aside, flexible filaments have been reportedly

produced by a few proteins with no remarks about loop formation, and this may indicate that filament flexibility need not imply end-joining [4, 39, 40, 60, 79].

The corollary to highly flexible filaments that are unable to end-join, is of course that end-joining can happen between stiff fibrils that cannot form loops. However, this can be expected to have a low probability of occurring between bundles of filaments, due to the low probability that the ends of the two bundles correctly align in experimentally observable time. Indeed, the findings of Lara et al. [60] appear to indicate that this is true of ovalbumin, where the thicker, straighter fibrils tend to fragment over time (becoming shorter), while the thinner, more flexible ones appear to become longer over time. This implies that filament flexibility, end-joining and bundle formation may be related in some way.

Chapter 8

Conclusions and outlook

In this thesis, I have shown how the system volume and protein concentration can be expected to alter the early-time kinetics of amyloid fibril self-assembly, and have explored the effects of lateral filament association and filament end-joining on the overall kinetics and filament length distribution. Each chapter makes distinct predictions regarding particular aspects of the self-assembly process. From these, two recurring themes emerge which raise important questions about the structure of amyloid fibrils, and whether experiments conducted *in vitro* are indicative of *in vivo* phenomena.

In chapter 4, I explored the waiting-time distributions produced by combinations of the processes of primary nucleation, polymerisation, and autocatalysis (by fragmentation or heterogeneous nucleation); for each model, I found that the distribution became narrower with increasing numbers of reactants (i.e. larger volume, higher protein concentration, or both). These revealed that the distributions of experimentally observed lag times are unlikely to be caused simply by fluctuations in the waiting times of these processes. Additionally, since these lag times are probably influenced by numerous factors rather than a single process, it may not be possible to determine which processes contribute without additional information; for example, examining the distributions produced at different volumes.

Most of the waiting-time distributions that I investigated, were obtained by neglecting depolymerisation. However, in the case where depolymerisation was included, it was found to be affected differently by the system volume, when compared with the other processes. Hence, further investigation of the effects of reversible polymerisation, particularly if it were length dependent, could provide additional insight into the experimentally observed lag-time variability.

In chapter 5, I demonstrated that "kinks" can appear in a plot of lag-time as a function of total protein concentration when the process dominating the lag phase changes at particular protein concentrations. Since these kinks become most apparent only when a wide range of protein concentrations are considered, I speculated that some of the plots reported in the literature [43, 87, 110, 111] may indicate their presence but not overtly enough to have been remarked on. Hence, the transition in these should become obvious if the experiments were repeated over a broader range of protein concentrations. However, the model also shows that under certain conditions — for example, particular system volumes or primary nucleation rates — these kinks may not be observed if the protein concentration at which the scaling exponent changes is not accessible experimentally.

Since this model was based on the analytic lag-time expression for the autocatalytic polymerisation model, it cannot address values of the lag-time scaling exponent below $\gamma = 1/2$, predicted by that model. However, $\gamma < 1/2$ has been reported for various proteins [18, 27, 53, 55, 74, 83, 94, 117] which indicates that an, as yet unidentified combination of processes must be affecting the lag phase of amyloid fibril formation, for some proteins, under certain conditions. However, these may still exhibit a change in scaling behaviour at particular protein concentrations, and this could be used as a means to identify the origin of these unexplained scaling exponent values.

The combined findings of chapters 4 and 5, demonstrate how crucial the system volume can be in determining which of primary nucleation or autocatalysis dominates the lag phase. At a volume comparable with a single biological cell, these results indicate that the lag time should be dominated by primary nucleation: compared with larger volumes, the average lag-time at the cellular scale is expected to be much longer and the variance between replicates is expected to be greater. Hence, confining high concentrations of protein to small

compartments within cells could be thought to have an evolutionary advantage, since it might help to mitigate the nucleation of any aggregate (not only amyloid fibrils), for similar reasons.

Given the way that the system size is expected to affect both the mean and variance in waiting time, experiments performed in volumes comparable with that of a single cell, should provide the best means to understand the early stages of amyloid fibril self-assembly. As such, further development of the microfluidics technique reported by Knowles et al. [56] could be of interest, particularly if it can be used to study effects in live cells. Indeed, the long average waiting-time for primary nucleation within a single cell might indicate why most amyloid diseases are associated with later life. However, the variability in nucleation time is unlikely to explain the variability in the age at which a disease is detectable, as many other factors (for example, the rate of cell damage) can be expected to influence this even after fibrils begin to assemble. Hence, microfluidics experiments of this kind present the best chance of resolving such questions.

Moving away from the effects of system size and the early-time kinetics: in chapters 6 and 7, I explored the consequences of lateral filament association and filament end-joining. While these were discussed in separate chapters, recent work by Lara et al. [60] may indicate that they are related. This raises important questions about different morphologies of amyloid fibrils, and the processes by which these form.

In chapter 6, I demonstrated that a modified version of the fragmentation model — in which fragmentation is arrested once a critical concentration of filaments is reached — was able to explain aspects of the experimental kinetics of bovine insulin which could not otherwise be captured by the fragmentation model. This modified model resulted in significantly higher populations of short filaments for total protein concentrations below the critical concentration, compared with those protein concentrations above it. Moreover, the model predicted that for total protein concentrations above the threshold, the highest concentration of short filaments occurred during the exponential growth phase, which coincides with the highest rate of cell damage reported [100] for amyloid beta. This implies that a similar transition may exist for amyloid beta as for bovine insulin; hence, it should be possible to find environmental conditions under which fibrils form, but

very little cell damage is observed. Clearly this could be of immense importance to the understanding and potential prevention of amyloid diseases.

In this model, I assumed that bundle formation by lateral filament association was responsible for the link between the critical filament concentration and the change in fragmentation rate. However, I only considered a coarse-grained implementation of this link: the same results would be expected if the fragmentation rate were affected by the critical concentration via another mechanism. Hence, in order to verify that bundle formation is indeed responsible for suppressing fragmentation, lateral filament association will need to be modelled explicitly, and the resulting bundle width compared with that obtained experimentally. However, such a model will probably require a better understanding of the bundle formation process; for example, whether the addition of new filaments to a bundle depends on the bundles size.

Finally, in chapter 7 I presented analytic solutions to models in which filaments are able to grow by monomer addition and end-joining. Using these I was able to obtain fitted values of the model parameters from experimental growth curves of ovalbumin, which, when simulated, produced kinetics very similar to those observed experimentally. The agreement of this model with the experimental results indicates that an equilibrium number of filaments exists for the system; thus, if the rate of fragmentation were artificially increased temporarily (perhaps through sonication), then the short filaments produced would be expected to reassemble into longer filaments over time.

The experimental results of a number of other proteins that exhibit similar growth curves and apparent end-joining [41, 51, 102], may also be explained by this model. If so, then filament end-joining could be a widely occurring and important process in the formation of amyloid fibrils. However, the analytic solutions that I presented, do not consider the origin of the high concentration of seeds necessary to produce the observed growth curves, nor do they take into account the formation of loops, which are clearly observed for some proteins and not others. Consequently, these solutions provide a useful and general understanding of some of the effects of end-joining, but the model clearly requires further development to address early-time behaviour and further understand loop formation.

A number of fibril forming proteins which have been linked with cell damage (for example Aβ and α-syn) also display kinetics similar to those of the end-joining model [4, 29, 30, 41, 51, 79, 98, 102]. While the medical implications of end-joining and loop formation are currently unclear, the end-joining mechanism will reduce the number of filament ends and shift the distribution of filament lengths towards longer filaments. This is in contrast to both fragmentation, which increases the number of filament ends and reduces their individual length, and bundling, which leaves the number of filament ends unchanged but reduces the exposed surface area. Hence, each of these mechanisms can be expected to exhibit differing degrees of damage to the cell, but which of them can be considered the most dangerous will only be determined through extensive *in vivo* studies.

While presented in separate chapters, it can be seen how lateral filament association and end-joining may affect one another. Individual filaments and thin bundles can be expected to have a higher flexibility than thicker bundles; thus, loop formation of the latter will be less likely. Indeed end-joining in general is likely to be harder for thicker bundles due to the reduced probability that the ends of all individual filaments in the bundle align appropriately. Similarly, if a threshold concentration of filaments is required for bundles to form, then end-joining may prevent their formation by keeping the number of filaments below this threshold (depending on the equilibrium value, N^*). Consequently, if the rate of lateral association or end-joining of filaments were affected by environmental factors such as the addition of salt, it might explain the change in physical appearance and kinetics exhibited by some proteins upon changing the experimental environment [35, 37, 60, 96].

The self-assembly properties of amyloid fibrils and other protein complexes have provided inspiration for the design of novel materials [36, 73, 109, 115] and also as a template or scaffold for the assembly of other materials [10, 88, 93]. In all such applications, the fibrils may be subject to sheer or tensile stresses, and so the structural properties of the fibrils will be very important. Hence, control of the fibril thickness is expected to be useful, as this can allow greater flexibility or resistance to breakage. Indeed, if bundle formation and end-joining are related, it may be beneficial to sacrifice strength in favour of the self-healing properties that thinner bundles may provide. While such advances in materials technology might not be expected for some time, identifying existing materials with potentially useful properties, is an important first step.

In this thesis I have made extensive use of "minimal models" in order to explore the effects of particular mechanisms. Such models are often used to explore the behaviour of biological systems, which are often intractably complicated. By neglecting the microscopic details of the processes thought to be involved in amyloid fibril self-assembly, I was able to infer how adjustment of the environmental conditions could determine which mechanism dominates the lag phase and also control the filament length distribution. These findings clearly have important implications for amyloid fibril self-assembly, but more experimental work is needed in order to verify the predictions of the model. Consequently, the questions raised by amyloid fibril self-assembly, and indeed most problems faced by science in the 21^{st} century, can only be addressed by a multi-faceted approach, with strong collaboration between experimental inquiry and theoretical modelling.

Appendix

A1 First passage times for reversible polymerisation using the backward Fokker-Plank equation

For a single particle performing a random walk on a one dimensional lattice, the time T at which it leaves an interval (a, b) can be obtained by making use of the backward Fokker-Plank equation [28]. This approach is general, but here I will consider only the case of one reflecting barrier at a and one absorbing barrier at b. Hence, for a particle starting at x, where $a \leq x \leq b$, the waiting time T is defined as the time at which the particles position is $x = b$.

The probability that the particle is still in the interval (a, b) at time t is given by

$$\int_a^b dx' p(x', t|x, 0) \equiv G(x, t) \tag{A1.1}$$

where $G(x, t)$ is an as yet unknown function. The condition $p(x', 0|x, 0) = \delta(x - x')$, along with the fact that a is reflective and b immediately absorbing, can be used to find the boundary conditions on G:

$$G(x, 0) = \begin{cases} 1 & \text{if } a \leq x \leq b \\ 0 & \text{else} \end{cases} \tag{A1.2}$$

$$G(b, t) = 0 \tag{A1.3}$$

$$\partial_x G(a, t) = 0 \tag{A1.4}$$

The probability $P(T \geq t)$ that the time at which the particle leaves is greater than t, is actually the same as equation A1.1, meaning that

$$p(x', t|x, 0) = p(x', 0|x, -t) \tag{A1.5}$$

and the dynamics are time reversible.

By writing the backward Fokker-Plank equation for $p(x', t|x, 0)$ it can be seen that $G(x, t)$ obeys

$$\partial_t G(x,t) = \partial_x \big[\omega(x,t)G(x,t)\big] + \frac{1}{2}\partial_x^2 \big[D(x,t)G(x,t)\big] \tag{A1.6}$$

Because the function $G(x, t)$ is also the probability that $T \geq t$, equation A1.6 can be used to obtain a solution to any moment of the exit-time probability distribution $P(T)$.

$$< T^n >= -\int_0^\infty t^{n-1} dG(x,t) = \int_0^\infty t^{n-1} G(x,t) dt \tag{A1.7}$$

A2 Waiting-time distributions for sequential Poisson processes with variable rates

The derivation of the distributions in §4.2.2, §4.2.3, and §4.2.4 all follow the same basic principle.

For a process with time dependent rate $\lambda(t)$ and Poisson distributed waiting-times, the probability that the k^{th} event in a sequence occurs in the interval $t \to t + dt$ is given by

$$P_k(t + dt) = P_k(t)\,(1 - \lambda(t)dt) + P_{k-1}(t + dt)\lambda(t)dt \tag{A2.1}$$

which, by the finite difference approximation, gives

$$\frac{dP_k(t)}{dt} = \lambda(t)\ (P_{k-1}(t) - P_k(t)) \tag{A2.2}$$

This can be solved iteratively by starting with the condition $P_0(0) = 1$ that no events have happened at $t = 0$:

$$P_k(t) = \lambda(t)\frac{\Lambda^{k-1}(t)}{(k-1)!}e^{-\Lambda(t)} \tag{A2.3}$$

where $\Lambda(\tau) = \int_0^\tau \lambda(t)dt$.

If λ is time dependent, then $\Lambda(t) \to \lambda t$ which recovers the Erlang distribution (equation 4.3). If, on the other hand, each step may have a different rate, then the probability of the k^{th} event occurring in time $t \to t + dt$ is

$$P_k(t + dt) = P_k(t)\,(1 - \lambda_{k,k+1}(t)dt) + P_{k-1}(t + dt)\lambda_{k-1,k}(t)dt \tag{A2.4}$$

where the rate of going from step $k - 1$ to step k is $\lambda_{k-1,k}$. This can also be solved iteratively, and converges to a closed from when at most one step has the different rate $\alpha \neq \lambda$:

$$P(\tau; k, \lambda, \alpha) = \frac{\alpha\lambda^k e^{-\alpha\tau}}{(\lambda - \alpha)^k}\left[1 - \frac{\Gamma(k, (\lambda - \alpha)\tau)}{\Gamma(k)}\right] \tag{A2.5}$$

This also recovers the Erlang distribution when $\alpha = \lambda$.

A3 The Van Kampen system size expansion

The Van Kampen approximation is a technique developed by van Kampen [105] for finding an approximate solution to a given master equation in the stochastic limit (i.e. far from the deterministic limit). This involves finding a small parameter, typically the inverse of the system size $1/V$, about which the equations can be expanded.

My collaborator, Juraj Szavitz-Nossan, identified that this approach could be used to find the waiting-time distribution $L(t)$ for the cooperative growth of a pair of coupled reactions ($\mathcal{N}(t)$ and $\mathcal{M}(t)$) to reach a threshold ($M_\phi = \mathcal{M}/N_\mathrm{A}V$), where the reactions depend on time-independent rate constants (ν and μ).

The pair of coupled reactions for the increase in the number of filaments $\mathcal{N}$ or their total monomer content $\mathcal{M}$ by a single filament or monomer, can be written as

$$\mathcal{N} \xrightarrow{\mathcal{M}\nu} \mathcal{N} + 1 \tag{A3.1}$$

$$\mathcal{M} \xrightarrow{\mathcal{N}\mu} \mathcal{M} + 1 \tag{A3.2}$$

and also in the form of a master equation

$$\frac{dP_{\mathcal{N},\mathcal{M}}}{dt} = \mathcal{M}\nu P_{\mathcal{N}-1,\mathcal{M}} + \mathcal{N}\mu P_{\mathcal{N},\mathcal{M}-1} - (\mathcal{M}\nu - \mathcal{N}\mu)P_{\mathcal{N},\mathcal{M}} \tag{A3.3}$$

In these, the small parameter will be $1/v = 1/N_\mathrm{A}V$, where V is the system volume and N_A Avogadro's constant. Using this small parameter, two new continuous random variables w and u can be defined:

$$w(t) = \frac{\mathcal{M}(t) - vM(t)}{\sqrt{v}} \qquad u(t) = \frac{\mathcal{N}(t) - vN(t)}{\sqrt{v}} \tag{A3.4}$$

where $M(t)$ and $N(t)$ will turn out to be the familiar concentrations of fibrillar protein and number of filaments, at time t. Expressing the left hand side of equation A3.3 in these continuous variables gives

$$\frac{d}{dt}P(\mathcal{M},\mathcal{N},t) = \frac{d}{dt}\tilde{P}(w,u,t) = \frac{\partial\tilde{P}}{\partial t} - \sqrt{v}\frac{dM(t)}{dt}\frac{\partial\tilde{P}}{\partial w} - \sqrt{v}\frac{dN(t)}{dt}\frac{\partial\tilde{P}}{\partial u} \tag{A3.5}$$

and so the entire master equation [1] becomes

$$\frac{\partial \tilde{P}(w,u,t)}{\partial t} = -\nu w\frac{\partial \tilde{P}}{\partial u} - \mu u\frac{\partial \tilde{P}}{\partial w} + \frac{1}{2}\nu M(t)\frac{\partial^2 \tilde{P}}{\partial u^2} + \frac{1}{2}\mu N(t)\frac{\partial^2 \tilde{P}}{\partial w^2} + \mathcal{O}\left(v^{-1/2}\right) \quad \text{(A3.6)}$$

which satisfies $\tilde{P}(w,u,0) = \delta(w)\delta(u)$, and where $M(t)$ and $N(t)$ also satisfy

$$\frac{dM(t)}{dt} = \mu N(t) \qquad \frac{dN(t)}{dt} = \nu M(t) \quad \text{(A3.7)}$$

$$M(0) = \frac{\mathcal{M}_0}{v} \qquad N(0) = \frac{\mathcal{N}_0}{v} \quad \text{(A3.8)}$$

These can be trivially solved to show that both $M(t)$ and $N(t)$ grow exponentially from their initial values $M(0)$ and $N(0)$; a number of approaches exist for finding $\partial\tilde{P}(w,u,t)$, but I will not present these steps as they do not add to the understanding of the model.

Using the probability distribution, the distribution of waiting times $L(t)$ for $M(t)$ to reach M_ϕ, can be found by considering the probability that in the interval $[t, t+dt]$, $M(t)$ passes M_ϕ, assuming that it had not passed it at t:

$$L(t)dt = \text{Prob}[M(t) > M_\phi, t+dt] - \text{Prob}[M(t) > M_\phi, t] \quad \text{(A3.9)}$$

for which

$$\text{Prob}[M(t) > M_\phi, t] = \int_{-\infty}^{\infty} du \int_{(M_\phi - M(t))\sqrt{v}}^{\infty} dw\, \tilde{P}(w,u,t) \quad \text{(A3.10)}$$

The solution to this, is the lag-time distribution

$$L(t) = \frac{(2F(t)M'(t) + (M_\phi - M(t))F'(t))\sqrt{v}}{2\sqrt{2\nu\pi F^3(t)}} e^{-\frac{(M_\phi - M(t))^2 v}{2\nu F(t)}} \quad \text{(A3.11)}$$

[1] Here, I also write the transition rates μ and ν in terms of w and u.

where $F(t)$ and $F'(t)$ are given by

$$F(t) = \frac{\mu}{3\nu}\left(\frac{M(0)}{\mu} + \frac{N(0)}{\nu}\right)[\cosh(\sqrt{\nu\mu}t) - 1]^2 + \\ + \frac{2\mu(N(0) + M(0))}{3\nu\sqrt{\nu\mu}}\sinh(\sqrt{\nu\mu}t)[\cosh(\sqrt{\nu\mu}t) - 1] + \\ + \frac{M(0)}{\nu}[\cosh(\sqrt{\nu\mu}t) - 1] + \frac{\mu N(0)}{\nu\sqrt{\nu\mu}}\sinh(\sqrt{\nu\mu}t)$$

and

$$F'(t) = \frac{2\mu}{3\nu}\left(N(0)\sqrt{\frac{\mu}{\nu}} + M(0)\sqrt{\frac{\nu}{\mu}}\right)[\cosh(\sqrt{\nu\mu}t) - 1]\sinh(\sqrt{\nu\mu}t) + \\ + \frac{2\mu(N(0) + M(0))}{3\nu}[2\cosh^2(\sqrt{\nu\mu}t) - \cosh(\sqrt{\nu\mu}t) - 1] + \\ + M(0)\sqrt{\frac{\mu}{\nu}}\sinh(\sqrt{\nu\mu}t) + \frac{\mu N(0)}{\nu}\cosh(\sqrt{\nu\mu}t)$$

It should be noted that this distribution does not give a unit integral over all t when v becomes "small". This indicates that there is a non-zero probability that the threshold will never be reached when the number of particles becomes small. The probability of this can be found from the complementary error function, which also allows the definition of "small" to be understood in terms of a critical volume v_c. This is given by

$$\lim_{t\to\infty}\text{Prob}[M(t) < M_\phi, t] \sim \frac{1}{2\sqrt{\pi}}\sqrt{\frac{v_c}{v}}e^{-v/v_c}, \quad v \gg v_c$$

where v_c is given by

$$v_c = \frac{\frac{2}{3}(\nu M(0) + \mu N(0)) + \frac{4}{3}\sqrt{\nu\mu}\,(M(0) + N(0))}{\left(\sqrt{\nu}M(0) + \sqrt{\mu}N(0)\right)^2}.$$

The mean and standard deviation of $L(t)$ can be found by approximating the distribution as a Gaussian, such that the mean lag-time is

$$\mu_\tau = \frac{1}{\sqrt{\mu\nu}}\ln\left(\frac{M_\phi + \sqrt{{M_\phi}^2 + N^2(0)\frac{\mu}{\nu} - M^2(0)}}{N(0)\sqrt{\frac{\mu}{\nu}} + M(0)}\right) \tag{A3.12}$$

and the variance is

$$\sigma_\tau^2 = \frac{\nu F(\mu_\tau)}{\mu^2 N^2(\mu_\tau) v} \tag{A3.13}$$

Bibliography

[1] Paolo Arosio, Marten Beeg, Lucrece Nicoud, and Massimo Morbidelli. Time evolution of amyloid fibril length distribution described by a population balance model. *Chemical Engineering Science*, 78:21–32, August 2012. ISSN 00092509.

[2] Aderonke Babajide, Robert Farber, Ivo L. Hofacker, Jeff Inman, Alan S. Lapedes, and Peter F. Stadler. Exploring protein sequence space using knowledge-based potentials. *Journal of Theoretical Biology*, 212(1):35–46, September 2001. ISSN 0022-5193.

[3] Andrew J. Baldwin, Spencer J. Anthony-Cahill, Tuomas P. J. Knowles, Guy Lippens, John Christodoulou, Paul D. Barker, and Christopher M. Dobson. Measurement of amyloid fibril length distributions by inclusion of rotational motion in solution NMR diffusion measurements. *Angewandte Chemie (International ed. in English)*, 47(18):3385–7, January 2008. ISSN 1521-3773.

[4] Jan Bieschke, Qinghai Zhang, Evan T. Powers, Richard A. Lerner, and Jeffery W. Kelly. Oxidative metabolites accelerate Alzheimer's amyloidogenesis by a two-step mechanism, eliminating the requirement for nucleation. *Biochemistry*, 44(13):4977–83, April 2005. ISSN 0006-2960.

[5] Katrina J. Binger, Chi L. L. Pham, Leanne M. Wilson, Michael F. Bailey, Lynne J. Lawrence, Peter Schuck, and Geoffrey J. Howlett. Apolipoprotein C-II amyloid fibrils assemble via a reversible pathway that includes fibril breaking and rejoining. *Journal of Molecular Biology*, 376(4):1116–29, March 2008. ISSN 1089-8638.

[6] Marilyn F. Bishop and Frank A. Ferrone. Kinetics of nucleation-controlled polymerization. A perturbation treatment for use with a secondary pathway. *Biophysical Journal*, 46(5):631–44, November 1984. ISSN 0006-3495.

[7] Monica Bucciantini, Giulia Calloni, Fabrizio Chiti, Lucia Formigli, Daniele Nosi, Christopher M. Dobson, and Massimo Stefani. Prefibrillar amyloid protein aggregates share common features of cytotoxicity. *The Journal of Biological Chemistry*, 279(30):31374–82, July 2004. ISSN 0021-9258.

[8] Marc A. Caporini, Vikram S. Bajaj, Mikhail Veshtort, Anthony Fitzpatrick, Cait E. MacPhee, Michele Vendruscolo, Christopher M. Dobson, and Robert G. Griffin. Accurate determination of interstrand distances and alignment in amyloid fibrils by magic angle spinning NMR. *The Journal of Physical Chemistry B*, 114(42):13555–61, October 2010. ISSN 1520-5207.

[9] Jacqui M. A. Carnall, Christopher A. Waudby, Ana M. Belenguer, Marc C. A. Stuart, Jérôme J.-P. Peyralans, and Sijbren Otto. Mechanosensitive self-replication driven by self-organization. *Science*, 327(5972):1502–6, March 2010. ISSN 1095-9203.

[10] Kevin J. Channon, Glyn L. Devlin, and Cait E. MacPhee. Efficient energy transfer within self-assembling peptide fibers: a route to light-harvesting nanomaterials. *Journal of the American Chemical Society*, 131(35):12520–1, September 2009. ISSN 1520-5126.

[11] Elio A. Cino, Mikko Karttunen, and Wing-Yiu Choy. Effects of molecular crowding on the dynamics of intrinsically disordered proteins. *PLoS ONE*, 7(11):e49876, January 2012. ISSN 1932-6203.

[12] Samuel I. A. Cohen, Michele Vendruscolo, Christopher M. Dobson, and Tuomas P. J. Knowles. Nucleated polymerization with secondary pathways. II. Determination of self-consistent solutions to growth processes described by non-linear master equations. *The Journal of Chemical Physics*, 135(6): 065106, August 2011. ISSN 1089-7690.

[13] Samuel I. A. Cohen, Michele Vendruscolo, Christopher M. Dobson, and Tuomas P. J. Knowles. Nucleated polymerization with secondary pathways. III. Equilibrium behavior and oligomer populations. *The Journal of Chemical Physics*, 135(6):065107, August 2011. ISSN 1089-7690.

[14] Samuel I. A. Cohen, Michele Vendruscolo, Christopher M. Dobson, and Tuomas P. J. Knowles. Nucleated polymerisation in the presence of pre-formed seed filaments. *International Journal of Molecular Sciences*, 12(9): 5844–52, January 2011. ISSN 1422-0067.

[15] Samuel I. A. Cohen, Michele Vendruscolo, Mark E. Welland, Christopher M. Dobson, Eugene M. Terentjev, and Tuomas P. J. Knowles. Nucleated polymerization with secondary pathways. I. Time evolution of the principal moments. *The Journal of Chemical Physics*, 135(6):065105, August 2011. ISSN 1089-7690.

[16] Samuel I. A. Cohen, Michele Vendruscolo, Christopher M. Dobson, and Tuomas P. J. Knowles. From macroscopic measurements to microscopic mechanisms of protein aggregation. *Journal of Molecular Biology*, 421(2-3):160–71, August 2012. ISSN 1089-8638.

[17] Samuel I. A. Cohen, Michele Vendruscolo, Christopher M. Dobson, and Tuomas P. J. Knowles. The Kinetics and Mechanisms of Amyloid Formation. In Daniel Erik Otzen, editor, *Amyloid Fibrils and Prefibrillar Aggregates*, pages 183–209. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, February 2013. ISBN 9783527654185.

[18] Sean R. Collins, Adam Douglass, Ronald D. Vale, and Jonathan S. Weissman. Mechanism of prion propagation: amyloid growth occurs by monomer addition. *PLoS Biology*, 2(10):e321, October 2004. ISSN 1545-7885.

[19] Kelly A. Conway. Acceleration of oligomerization, not fibrillization, is a shared property of both α-synuclein mutations linked to early-onset Parkinson's disease: Implications for pathogenesis and therapy. *Proceedings of the National Academy of Science*, 97(2):571–576, January 2000. ISSN 00278424.

[20] Christopher M. Dobson. Protein misfolding, evolution and disease. *Trends in Biochemical Sciences*, 24(9):329–332, September 1999. ISSN 09680004.

[21] Christopher M. Dobson. Protein folding and misfolding. *Nature*, 426(6968): 884–90, December 2003. ISSN 1476-4687.

[22] Masao Doi. *The Theory of Polymer Dynamics*. Oxford University Press, 1988. ISBN 0198520336.

[23] M. Fandrich. The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *The EMBO Journal*, 21(21): 5682–5690, November 2002. ISSN 14602075.

[24] Frank A. Ferrone. Analysis of protein aggregation kinetics. *Methods in Enzymology*, 309(1):256–74, January 1999. ISSN 0076-6879.

[25] Frank A. Ferrone, James Hofrichter, Helen R. Sunshine, and William A. Eaton. Kinetic studies on photolysis-induced gelation of sickle cell hemoglobin suggest a new mechanism. *Biophysical Journal*, 32(1):361–80, October 1980. ISSN 0006-3495.

[26] Frank A. Ferrone, James Hofrichter, and William A. Eaton. Kinetics of sickle hemoglobin polymerization. *Journal of Molecular Biology*, 183(4): 611–631, June 1985. ISSN 00222836.

[27] Vito Foderà, Fabio Librizzi, Minna Groenning, Marco van de Weert, and Maurizio Leone. Secondary nucleation and accessible surface in insulin amyloid fibril formation. *The Journal of Physical Chemistry B*, 112(12): 3853–8, March 2008. ISSN 1520-6106.

[28] Crispin W. Gardiner. *Handbook of Stochastic Methods: For Physics, Chemistry and the Natural Sciences*. Springer, November 1996. ISBN 3540616349.

[29] Lise Giehm and Daniel E. Otzen. Strategies to increase the reproducibility of protein fibrillization in plate reader assays. *Analytical Biochemistry*, 400 (2):270–81, May 2010. ISSN 1096-0309.

[30] Lise Giehm, Cristiano Luis Pinto Oliveira, Gunna Christiansen, Jan Skov Pedersen, and Daniel E. Otzen. SDS-induced fibrillation of α-synuclein: an alternative fibrillation pathway. *Journal of Molecular Biology*, 401(1): 115–33, August 2010. ISSN 1089-8638.

[31] Daniel T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, December 1976. ISSN 00219991.

[32] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, December 1977. ISSN 0022-3654.

[33] Charles G. Glabe and Rakez Kayed. Common structure and toxic function of amyloid oligomers implies a common mechanism of pathogenesis. *Neurology*, 66(2 Suppl 1):S74–8, January 2006. ISSN 1526-632X.

[34] Walraj S. Gosal, Allan H. Clark, and Simon B. Ross-Murphy. Fibrillar β-lactoglobulin gels: Part 1. Fibril formation and structure. *Biomacromolecules*, 5(6):2408–19, 2004. ISSN 1525-7797.

[35] Walraj S. Gosal, Isobel J. Morten, Eric W. Hewitt, D. Alastair Smith, Neil H. Thomson, and Sheena E. Radford. Competing pathways determine fibril morphology in the self-assembly of β_2-microglobulin into amyloid. *Journal of Molecular Biology*, 351(4):850–64, August 2005. ISSN 0022-2836.

[36] Sally L. Gras. Amyloid Fibrils: From Disease to Design. New Biomaterial Applications for Self-Assembling Cross-β Fibrils. *Australian Journal of Chemistry*, 60(5):333, 2007. ISSN 0004-9425.

[37] Michael D. W. Griffin, Melva L. Y. Mok, Leanne M. Wilson, Chi L. L. Pham, Lynne J. Waddington, Matthew A. Perugini, and Geoffrey J. Howlett. Phospholipid interaction induces molecular-level polymorphism in apolipoprotein C-II amyloid fibrils via alternative assembly pathways. *Journal of Molecular Biology*, 375(1):240–56, January 2008. ISSN 1089-8638.

[38] James D. Harper and Peter T. Lansbury, Jr. Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annual Review of Biochemistry*, 66:385–407, January 1997. ISSN 0066-4154.

[39] James D. Harper, Charles M. Lieber, and Peter T. Lansbury, Jr. Atomic force microscopic imaging of seeded fibril formation and fibril branching by the Alzheimer's disease amyloid-β protein. *Chemistry & Biology*, 4(12): 951–959, December 1997. ISSN 10745521.

[40] James D. Harper, Stanislaus S. Wong, Charles M. Lieber, and Peter T. Lansbury, Jr. Assembly of Aβ amyloid protofibrils: an in vitro model for a possible early event in Alzheimer's disease. *Biochemistry*, 38(28):8972–80, July 1999. ISSN 0006-2960.

[41] Danny M. Hatters, Cait E. MacPhee, Lynne J. Lawrence, William H. Sawyer, and Geoffrey J. Howlett. Human apolipoprotein C-II forms twisted amyloid ribbons and closed loops. *Biochemistry*, 39(28):8276–83, July 2000. ISSN 0006-2960.

[42] Danny M. Hatters, Christopher A. MacRaild, Rob Daniels, Walraj S. Gosal, Neil H. Thomson, Jonathan A. Jones, Jason J. Davis, Cait E. MacPhee, Christopher M. Dobson, and Geoffrey J. Howlett. The circularization of amyloid fibrils formed by apolipoprotein C-II. *Biophysical Journal*, 85(6): 3979–90, December 2003. ISSN 0006-3495.

[43] Erik Hellstrand, Barry Boland, Dominic M. Walsh, and Sara Linse. Amyloid β-protein aggregation produces highly reproducible kinetic data and occurs by a two-phase process. *ACS Chemical Neuroscience*, 1(1):13–8, January 2010. ISSN 1948-7193.

[44] Terrell L. Hill. Length dependence of rate constants for end-to-end association and dissociation of equilibrium linear aggregates. *Biophysical Journal*, 44(2):285–8, November 1983. ISSN 0006-3495.

[45] James Hofrichter. Kinetics of sickle hemoglobin polymerization: III. Nucleation rates determined from stochastic fluctuations in polymerization progress curves. *Journal of Molecular Biology*, 189(3):553–571, June 1986. ISSN 00222836.

[46] Liu Hong and Wen-An Yong. Simple moment-closure model for the self-assembly of breakable amyloid filaments. *Biophysical Journal*, 104(3):533–40, March 2013. ISSN 1542-0086.

[47] Liu Hong, Xianghong Qi, and Yang Zhang. Dissecting the kinetic process of amyloid fiber formation through asymptotic analysis. *The Journal of Physical Chemistry B*, 116(23):6611–7, June 2012. ISSN 1520-5207.

[48] Peter Hortschansky, Volker Schroeckh, Tony Christopeit, Giorgia Zandomeneghi, and Marcus Fändrich. The aggregation kinetics of Alzheimer's β-amyloid peptide is controlled by stochastic nucleation. *Protein Science*, 14(7):1753–9, July 2005. ISSN 0961-8368.

[49] Ralf Jansen, Wojciech Dzwolak, and Roland Winter. Amyloidogenic self-assembly of insulin aggregates probed by high resolution atomic force microscopy. *Biophysical Journal*, 88(2):1344–53, March 2005. ISSN 0006-3495.

[50] Richard A. L. Jones. *Soft condensed matter*. OUP Oxford, 2002. ISBN 0198505892.

[51] Josué Juárez, Pablo Taboada, and Víctor Mosquera. Existence of different structural intermediates on the fibrillation pathway of human serum albumin. *Biophysical Journal*, 96(6):2353–70, March 2009. ISSN 1542-0086.

[52] Rakez Kayed, Elizabeth Head, Jennifer L. Thompson, Theresa M. McIntire, Saskia C. Milton, Carl W. Cotman, and Charles G. Glabe. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science*, 300(5618):486–9, April 2003. ISSN 1095-9203.

[53] Jeffery W. Kelly. Mechanisms of amyloidogenesis. *Nature structural biology*, 7(10):824–6, October 2000. ISSN 1072-8368.

[54] Robert Kisilevsky. Review: amyloidogenesis — unquestioned answers and unanswered questions. *Journal of Structural Biology*, 130(2-3):99–108, June 2000. ISSN 1047-8477.

[55] Tuomas P. J. Knowles, Christopher A. Waudby, Glyn L. Devlin, Samuel I. A. Cohen, Adriano Aguzzi, Michele Vendruscolo, Eugene M. Terentjev, Mark E. Welland, and Christopher M. Dobson. An analytical solution to the kinetics of breakable filament assembly. *Science*, 326(5959):1533–7, December 2009. ISSN 1095-9203.

[56] Tuomas P. J. Knowles, Duncan A. White, Adam R. Abate, Jeremy J. Agresti, Samuel I. A. Cohen, Ralph A. Sperling, Erwin J. de Genst, Christopher M. Dobson, and David A. Weitz. Observation of spatial propagation of amyloid assembly from single nuclei. *Proceedings of the National Academy of Science*, 108(36):14746–51, September 2011. ISSN 1091-6490.

[57] Kleanthes Koniaris and Murugappan Muthukumar. Self-entanglement in ring polymers. *The Journal of Chemical Physics*, 95(4):2873, 1991. ISSN 00219606.

[58] Kay C. Kunes, Daniel L. Cox, and Rajiv R. P. Singh. One-dimensional model of yeast prion aggregation. *Physical Review E*, 72(5):051915, November 2005. ISSN 1539-3755.

[59] Thomas G. Kurtz. The Relationship between Stochastic and Deterministic Models for Chemical Reactions. *The Journal of Chemical Physics*, 57(7): 2976, 1972. ISSN 00219606.

[60] Cecile Lara, Simon Gourdin-Bertin, Jozef Adamcik, Sreenath Bolisetty, and Raffaele Mezzenga. Self-assembly of ovalbumin into amyloid and non-amyloid fibrils. *Biomacromolecules*, 13(12):4213–21, December 2012. ISSN 1526-4602.

[61] Hilal A. Lashuel, Steven R. Labrenz, Linda Woo, Louise C. Serpell, and Jeffery W. Kelly. Protofilaments, filaments, ribbons, and fibrils from peptidomimetic self-assembly: implications for amyloid fibril formation and materials science. *Journal of the American Chemical Society*, 122(22):5262–77, June 2000. ISSN 0002-7863.

[62] Sylvain Lesné, Ming Teng Koh, Linda Kotilinek, Rakez Kayed, Charles G. Glabe, Austin Yang, Michela Gallagher, and Karen H. Ashe. A specific amyloid-β protein assembly in the brain impairs memory. *Nature*, 440 (7082):352–7, March 2006. ISSN 1476-4687.

[63] Yi Li and Christopher J. Roberts. Lumry-Eyring nucleated-polymerization model of protein aggregation kinetics. 2. Competing growth via condensation and chain polymerization. *The Journal of Physical Chemistry B*, 113(19):7020–32, May 2009. ISSN 1520-6106.

[64] Aleksey Lomakin, David B. Teplow, Daniel A. Kirschner, and George B. Benedek. Kinetic theory of fibrillogenesis of amyloid β-protein. *Proceedings of the National Academy of Science*, 94(15):7942–7, July 1997. ISSN 0027-8424.

[65] Qian Ma, Jun-Bao Fan, Zheng Zhou, Bing-Rui Zhou, Sheng-Rong Meng, Ji-Ying Hu, Jie Chen, and Yi Liang. The contrasting effect of macromolecular crowding on amyloid fibril formation. *PLoS ONE*, 7(4):e36288, January 2012. ISSN 1932-6203.

[66] Mauro Manno, Manno Mauro, Emanuela Fabiola Craparo, Alessandro Podestà, Donatella Bulone, Rita Carrotta, Vincenzo Martorana, Guido Tiana, and Pier Luigi San Biagio. Kinetics of different processes in human insulin amyloid formation. *Journal of Molecular Biology*, 366(1):258–74, March 2007. ISSN 0022-2836.

[67] Ivo Cristiano Martins, Inna Kuperstein, Hannah Wilkinson, Elke Maes, Mieke Vanbrabant, Wim Jonckheere, Patrick van Gelder, Dieter Hartmann, Rudi D'Hooge, Bart de Strooper, Joost Schymkowitz, and Frederic Rousseau. Lipids revert inert Abeta amyloid fibrils to neurotoxic protofibrils that affect learning in mice. *The EMBO Journal*, 27(1):224–33, January 2008. ISSN 1460-2075.

[68] Graham McNeill. *A Thousand Sons*. Games Workshop, 2010. ISBN 978-1844168088.

[69] Pierre-Emmanuel Milhiet, Daisuke Yamamoto, Olivia Berthoumieu, Patrice Dosset, Christian Le Grimellec, Jean-Michel Verdier, Stéphane Marchal, and Toshio Ando. Deciphering the structure, growth and assembly of amyloid-like fibrils using high-speed atomic force microscopy. *PLoS ONE*, 5(10):e13240, January 2010. ISSN 1932-6203.

[70] Yee-Foong Mok, Timothy M. Ryan, Shuo Yang, Danny M. Hatters, Geoffrey J. Howlett, and Michael D. W. Griffin. Sedimentation velocity analysis of amyloid oligomers and fibrils using fluorescence detection. *Methods*, 54(1):67–75, May 2011. ISSN 1095-9130.

[71] Aimee M. Morris, Murielle A. Watzky, Jeffrey N. Agar, and Richard G. Finke. Fitting neurological protein aggregation kinetic data via a 2-step, minimal/"Ockham's razor" model: the Finke-Watzky mechanism of nucleation followed by autocatalytic surface growth. *Biochemistry*, 47(8): 2413–27, March 2008. ISSN 0006-2960.

[72] Aimee M. Morris, Murielle A. Watzky, and Richard G. Finke. Protein aggregation kinetics, mechanism, and curve-fitting: a review of the literature. *Biochimica et Biophysica Acta*, 1794(3):375–97, March 2009. ISSN 0006-3002.

[73] Ryan J. Morris and Cait E. MacPhee. Amyloid Protein Biomaterials. *Encyclopedia of Biophysics*, pages 76–81, 2013. ISSN 364216711X.

[74] Ryan J. Morris, Kym Eden, Reuben Yarwood, Line Jourdain, Rosalind J. Allen, and Cait E. Macphee. Mechanistic and environmental control of the prevalence and lifetime of amyloid oligomers. *Nature communications*, 4 (May):1891, January 2013. ISSN 2041-1723.

[75] Larissa A. Munishkina and Anthony L. Fink. Fluorescence as a method to reveal structures and membrane-interactions of amyloidogenic proteins. *Biochimica et Biophysica Acta*, 1768(8):1862–85, August 2007. ISSN 0006-3002.

[76] Larissa A. Munishkina, Atta Ahmad, Anthony L. Fink, and Vladimir N. Uversky. Guiding protein aggregation with macromolecular crowding. *Biochemistry*, 47(34):8993–9006, August 2008. ISSN 1520-4995.

[77] David Nelson and Michael Cox. *Lehninger Principles of Biochemistry: Fourth Edition*. W.H. Freeman and Co., 2005. ISBN 978-0716762652.

[78] Rebecca Nelson and David Eisenberg. Structural models of amyloid-like fibrils. *Advances in Protein Chemistry*, 73(06):235–82, January 2006. ISSN 0065-3233.

[79] Michael R. Nichols, Melissa A. Moss, Dana Kim Reed, Stephanie Cratic-McDaniel, Jan H. Hoh, and Terrone L. Rosenberry. Amyloid-β protofibrils differ from amyloid-β aggregates induced in dilute hexafluoroisopropanol in stability and morphology. *The Journal of Biological Chemistry*, 280(4): 2471–80, January 2005. ISSN 0021-9258.

[80] Brian O'Nuallain, Darragh B. Freir, Andrew J. Nicoll, Emmanuel Risse, Neil Ferguson, Caroline E. Herron, John Collinge, and Dominic M. Walsh. Amyloid β-protein dimers rapidly form stable synaptotoxic protofibrils. *The Journal of Neuroscience*, 30(43):14411–9, October 2010. ISSN 1529-2401.

[81] Fumio Oosawa and Sho Asakura. *Thermodynamics of the polymerization of protein*. Academic Press Inc, 1975. ISBN 978-0125270502.

[82] Fumio Oosawa and Michiki Kasai. A theory of linear and helical aggregations of macromolecules. *Journal of Molecular Biology*, 4(1):10–21, January 1962. ISSN 00222836.

[83] Shae B. Padrick and Andrew D. Miranker. Islet amyloid: phase partitioning and secondary nucleation are central to the mechanism of fibrillogenesis. *Biochemistry*, 41(14):4694–703, April 2002. ISSN 0006-2960.

[84] Leonard F. Pease, Mirco Sorci, Suvajyoti Guha, De-Hao Tsai, Michael R. Zachariah, Michael J. Tarlov, and Georges Belfort. Probing the nucleus model for oligomer formation during insulin amyloid fibrillogenesis. *Biophysical Journal*, 99(12):3979–85, December 2010. ISSN 1542-0086.

[85] Evan T. Powers and David L. Powers. The kinetics of nucleated polymerizations at high concentrations: amyloid fibril formation near and above the "supercritical concentration". *Biophysical Journal*, 91(1):122–32, July 2006. ISSN 0006-3495.

[86] Sheena E. Radford and Christopher M. Dobson. From Computer Simulations to Human Disease. *Cell*, 97(3):291–298, April 1999. ISSN 00928674.

[87] Gayathri Ramachandran and Jayant B. Udgaonkar. Evidence for the existence of a secondary pathway for fibril growth during the aggregation of tau. *Journal of Molecular Biology*, 421(2-3):296–314, August 2012. ISSN 1089-8638.

[88] Meital Reches and Ehud Gazit. Casting metal nanowires within discrete self-assembled peptide nanotubes. *Science*, 300(5619):625–7, April 2003. ISSN 1095-9203.

[89] Sidney Redner. *A Guide To First-Passage Processes*. Cambridge University Press, Cambridge, 2001. ISBN 9780511606014.

[90] Elizabeth Rhoades, Jayant Agarwal, and Ari Gafni. Aggregation of an amyloidogenic fragment of human islet amyloid polypeptide. *Biochimica et Biophysica Acta*, 1476(2):230–8, March 2000. ISSN 0006-3002.

[91] Timothy M. Ryan, Chai Lean Teoh, Michael D. W. Griffin, Michael F. Bailey, Peter Schuck, and Geoffrey J. Howlett. Phospholipids enhance nucleation but not elongation of apolipoprotein C-II amyloid fibrils. *Journal of Molecular Biology*, 399(5):731–40, June 2010. ISSN 1089-8638.

[92] Timothy M. Ryan, Michael D. W. Griffin, Chai Lean Teoh, Jingkai Ooi, and Geoffrey J. Howlett. High-affinity amphipathic modulators of amyloid fibril nucleation and elongation. *Journal of Molecular Biology*, 406(3):416–29, March 2011. ISSN 1089-8638.

[93] Thomas Scheibel, Raghuveer Parthasarathy, George J. Sawicki, Xiao-Min Lin, Heinrich Jaeger, and Susan L. Lindquist. Conducting nanowires built by controlled self-assembly of amyloid fibers and selective metal deposition. *Proceedings of the National Academy of Science*, 100(8):4527–32, April 2003. ISSN 0027-8424.

[94] Tricia R. Serio, Anil G. Cashikar, Anthony S. Kowal, George J. Sawicki, Jahan J. Moslehi, Louise Serpell, Morton F. Arnsdorf, and Susan L. Lindquist. Nucleated Conformational Conversion and the Replication of Conformational Information by a Prion Determinant. *Science*, 289(5483): 1317–1321, August 2000. ISSN 00368075.

[95] Ganesh M. Shankar, Shaomin Li, Tapan H. Mehta, Amaya Garcia-Munoz, Nina E. Shepardson, Imelda Smith, Francesca M. Brett, Michael A. Farrell, Michael J. Rowan, Cynthia A. Lemere, Ciaran M. Regan, Dominic M. Walsh, Bernardo L. Sabatini, and Dennis J. Selkoe. Amyloid-β protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory. *Nature medicine*, 14(8):837–42, August 2008. ISSN 1546-170X.

[96] Andrew M. Smith, Thomas R. Jahn, Alison E. Ashcroft, and Sheena E. Radford. Direct observation of oligomeric species formed in the early stages of amyloid fibril formation using electrospray ionisation mass spectrometry. *Journal of Molecular Biology*, 364(1):9–19, November 2006. ISSN 0022-2836.

[97] Mirco Sorci, Robert A. Grassucci, Ingrid Hahn, Joachim Frank, and Georges Belfort. Time-dependent insulin oligomer reaction pathway prior to fibril formation: cooling and seeding. *Proteins*, 77(1):62–73, October 2009. ISSN 1097-0134.

[98] Andrea Stirpe, Manuela Pantusa, Bruno Rizzuti, Luigi Sportelli, Rosa Bartucci, and Rita Guzzi. Early stage aggregation of human serum albumin in the presence of metal ions. *International Journal of Biological Macromolecules*, 49(3):337–42, October 2011. ISSN 1879-0003.

[99] Margaret Sunde, Louise C. Serpell, Mark Bartlam, Paul E. Fraser, Mark B. Pepys, and Colin C. F. Blake. Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *Journal of Molecular Biology*, 273(3):729–39, October 1997. ISSN 0022-2836.

[100] Bruce M. Taylor, Ronald W. Sarver, Gregory Fici, Roger A. Poorman, Barry S. Lutzke, Antonio Molinari, Thomas Kawabe, Karl Kappenman, Allen E. Buhl, and Dennis E. Epps. Spontaneous aggregation and cytotoxicity of the β-amyloid Aβ^{1-40}: a kinetic model. *Journal of Protein Chemistry*, 22(1):31–40, January 2003. ISSN 0277-8033.

[101] Chai Lean Teoh, Chi L. L. Pham, Nevena Todorova, Andrew Hung, Craig N. Lincoln, Emma Lees, Yuen Han Lam, Katrina J. Binger, Neil H. Thomson, Sheena E. Radford, Trevor A. Smith, Shirley A. Müller, Andreas Engel, Michael D. W. Griffin, Irene Yarovsky, Paul R. Gooley, and Geoffrey J. Howlett. A structural model for apolipoprotein C-II amyloid fibrils: experimental characterization and molecular dynamics simulations. *Journal of Molecular Biology*, 405(5):1246–66, March 2011. ISSN 1089-8638.

[102] David C. Thorn, Heath Ecroyd, Margaret Sunde, Stephen Poon, and John A. Carver. Amyloid fibril formation by bovine milk α_{s2}-casein occurs under physiological conditions yet is prevented by its natural counterpart, α_{s1}-casein. *Biochemistry*, 47(12):3926–36, March 2008. ISSN 0006-2960.

[103] Darwin Thusius. Mechanism of bovine liver glutamate dehydrogenase self-assembly. *Journal of Molecular Biology*, 94(3):367–383, May 1975. ISSN 00222836.

[104] Darwin Thusius, Philippe Dessen, and Jean-Marc Jallon. Mechanism of bovine liver glutamate dehydrogenase self-association. *Journal of Molecular Biology*, 92(3):413–432, March 1975. ISSN 00222836.

[105] Nicolaas G. van Kampen. A power series expansion of the master equation. *Canadian Journal of Physics*, 39(4):551–567, April 1961. ISSN 0008-4204.

[106] Albrecht Wegner and Paula Savko. Fragmentation of actin filaments. *Biochemistry*, 21(8):1909–13, April 1982. ISSN 0006-2960.

[107] Mireille Weijers, Kerensa Broersen, Peter A. Barneveld, Martien A. Cohen Stuart, Rob J. Hamer, Harmen H. J. de Jongh, and Ronald W. Visschers. Net charge affects morphology and visual properties of ovalbumin aggregates. *Biomacromolecules*, 9(11):3165–72, November 2008. ISSN 1526-4602.

[108] Duncan A. White, Alexander K. Buell, Tuomas P. J. Knowles, Mark E. Welland, and Christopher M. Dobson. Protein aggregation in crowded environments. *Journal of the American Chemical Society*, 132(14):5170–5, April 2010. ISSN 1520-5126.

[109] Derek N. Woolfson and Maxim G. Ryadnov. Peptide-based fibrous biomaterials: Some things old, new and borrowed. *Current Opinion in Chemical Biology*, 10(6):559–67, December 2006. ISSN 1367-5931.

[110] Caroline F. Wright, Sarah A. Teichmann, Jane Clarke, and Christopher M. Dobson. The importance of sequence diversity in the aggregation and evolution of proteins. *Nature*, 438(7069):878–81, December 2005. ISSN 1476-4687.

[111] Wei-Feng Xue, Steve W. Homans, and Sheena E. Radford. Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proceedings of the National Academy of Science*, 105(26):8926–31, July 2008. ISSN 1091-6490.

[112] Wei-Feng Xue, Andrew L. Hellewell, Walraj S. Gosal, Steve W. Homans, Eric W. Hewitt, and Sheena E. Radford. Fibril fragmentation enhances amyloid cytotoxicity. *The Journal of Biological Chemistry*, 284(49):34272–82, December 2009. ISSN 1083-351X.

[113] Takahiro Yamaguchi, Hisashi Yagi, Yuji Goto, Katsumi Matsuzaki, and Masaru Hoshino. A disulfide-linked amyloid-β peptide dimer forms a protofibril-like oligomer through a distinct pathway from amyloid fibril formation. *Biochemistry*, 49(33):7100–7, August 2010. ISSN 1520-4995.

[114] Shuo Yang, Michael D. W. Griffin, Katrina J. Binger, Peter Schuck, and Geoffrey J. Howlett. An equilibrium model for linear and closed-loop amyloid fibril formation. *Journal of Molecular Biology*, 421(2-3):364–77, August 2012. ISSN 1089-8638.

[115] Shuguang Zhang. Fabrication of novel biomaterials through molecular self-assembly. *Nature biotechnology*, 21(10):1171–8, October 2003. ISSN 1087-0156.

[116] Zheng Zhou, Jun-Bao Fan, Hai-Li Zhu, Frank Shewmaker, Xu Yan, Xi Chen, Jie Chen, Geng-Fu Xiao, Lin Guo, and Yi Liang. Crowded cell-like environment accelerates the nucleation step of amyloidogenic protein misfolding. *The Journal of Biological Chemistry*, 284(44):30148–58, October 2009. ISSN 1083-351X.

[117] Li Zhu, Xu-Jia Zhang, Ling-Yun Wang, Jun-Mei Zhou, and Sarah Perrett. Relationship Between Stability of Folding Intermediates and Amyloid Formation for the Yeast Prion Ure2p: A Quantitative Analysis of the Effects of pH and Buffer System. *Journal of Molecular Biology*, 328(1):235–254, April 2003. ISSN 00222836.

Publications

Ryan J. Morris, Kym Eden, Reuben Yarwood, Line Jourdain, Rosalind J. Allen, and Cait E. Macphee. Mechanistic and environmental control of the prevalence and lifetime of amyloid oligomers. *Nature communications*, 4(May):1891, January 2013. ISSN 2041-1723